



Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh

Citation

Scanlon, Thomas M. 2002. Reasons, responsibility, and reliance: Replies to Wallace, Dworkin, and Deigh. *Ethics* 112, no. 3: 507-528.

Published Version

<http://dx.doi.org/10.1086/339221>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3708348>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Reasons, Responsibility, and Reliance: Replies to Wallace, Dworkin, and Deigh

T. M. Scanlon

I am very grateful to John Deigh, Jerry Dworkin, and Jay Wallace for their perceptive and challenging comments on my book. I have learned a lot from thinking about them. I will begin by responding to Wallace and Dworkin, whose comments deal with general issues in moral theory, and then turn to Deigh's critique of my account of promises.

REPLY TO WALLACE

Reasons, Desires, and Motivation

In chapter 1 of *What We Owe to Each Other*, I make several claims about reasons and motivation which are perhaps not as clear as they might have been, in part because I find the notion of motivation itself rather unclear. First, I claim that if an agent is rational, then the fact that she sees some consideration as a reason for acting in a certain way can itself explain her forming an intention to act in that way, and her so acting. There are conditions that must be present in order for an agent's taking something to be a reason to lead to the formation of an intention and to subsequent action. It is necessary, for example, that the agent not be rendered unconscious, physically interfered with, or distracted by other thoughts. But what is not necessary, in my view, is the presence of some further "motivating" state of desire, over and above the judgment-sensitive attitude of taking some consideration to be a reason. Of course, that attitude itself, just insofar as it is capable of explaining action, may be counted as a desire in the broad sense in which that term is often used. What is essential to my claim, however, is that no additional motivational state is required: the fact that the state I call seeing something as a reason has cognitive content does not debar it from serving as the motivation for an action.

A second, broader claim that I make is that what we commonly call desires have just this cognitive component: they involve seeing some consideration as a reason. This claim seems to me to be supported by

reflection on the phenomenology of desires. The element of seeing something as in some way worth doing, or worth bringing about, is what differentiates desires from mere urges. If this claim is correct, then the motivational efficacy of desires can be understood as an instance of the phenomenon of being moved by seeing something as a reason. I am not certain I would follow Wallace in calling this form of motivation a "more basic mechanism" (p. 436). I am claiming that the idea of seeing something as a reason is already presupposed by our ordinary notion of desire, rather than a notion that is to be explained in terms of it. So it may be "more basic" in that sense. But there may be other senses of "more basic" in which the reverse is true.

Cats, dogs, and infants are not capable of assessing reasons and are therefore not capable of acting on an intention (understood as a judgment-sensitive attitude). They do have appetitive states, which move them to goal-directed action. Perhaps these states are more basic (more primitive, less developed) than judgment-sensitive attitudes. It is plausible to think that we adult humans are also sometimes moved in ways that are best understood in this fashion. We may not be best understood as acting for a reason when we reflexively pull away from something painful, for example, or when we are semiconscious, due to sleep or anesthesia, and struggle to get air or to free ourselves from some constraint. What I resist, however, is the suggestion that most of our motivation is of this "more basic" kind, or that motivation by reasons needs to be explained in terms of it. It seems to me, on the contrary, that in our case the desires that move us to action almost always involve our seeing something (other than that desire itself) as counting in favor of so acting. If this is correct, then it follows that the desires that commonly move us differ in an important way from the states that move animals and infants. This may initially have an implausible ring about it, but I think that on reflection it is the description that best fits the facts.

Irrationality and Akrasia

In my view, irrationality occurs when a person holds a judgment-sensitive attitude that is incompatible with his or her own assessment of the reasons that bear on having that attitude. This happens, for example, when a person judges there to be compelling reason to adopt a certain belief but fails to come to have this belief, or judges there to be compelling reason to abandon an intention but continues to have that intention.

If we were perfectly rational then this would never happen; our attitudes would always accord with our judgments. It is a familiar fact—regrettable but not in itself puzzling—that this is not the case. Nonetheless, irrational attitudes can seem puzzling, even impossible, if the attitudes in question are understood in a way that involves the relevant underlying judgments so directly that these attitudes cannot be held with-

out these judgments being present as well. For example, if intending to act in a certain way involved judging that this was the way one had most reason to act, then one could not act irrationally. On this account it would be impossible to intentionally drink a fifth glass of wine while believing that one should not do this, because forming the intention to have the drink would involve modifying this underlying judgment.

As Davidson has convincingly argued, however, intentional action need not be understood in this way: one can decide to act in a certain way, for a certain reason, without judging that to be, on balance, the way one has most reason to act.¹ One can decide to do something because it would be fun, knowing all the while that there are compelling reasons not to do it. In order for this to happen, a person must fail to give this particular reason the normative force that he or she takes it to have. In this respect, such a person will fail to be "clear-eyed." But accepting this view of *akratic* action would amount to skepticism about the possibility of *akrasia* only if such a failure to be "clear-eyed" involved going back on one's assessment of the reasons against one's course of action. Some cases of "acting against one's better judgment" may be of this kind (that is, they may involve modifying this better judgment at least temporarily), but not all of them do, and my description of *akrasia* does not require this. As Wallace points out with his example of strength of will, people vary in the degree to which their considered normative assessment of various considerations is what governs their conduct. But I think this is best seen simply as a fact about the varying degrees to which people's actions match their judgments. I do not think that clarity is advanced by introducing an executive capacity, the will, whose strength or weakness is supposed to explain this variation.

Akrasia of belief is more puzzling than *akrasia* of action, and may seem to be impossible, because it is natural to understand belief that *p* as necessarily involving, or perhaps even consisting primarily in, taking there to be sufficient reason to think that *p*. But belief is not best understood in this way. Believing that *p* involves things other than accepting the judgment that this belief is warranted. It also includes such things as a readiness to take *p* as a premise in argument and to take other considerations to be reasons that would not be reasons unless *p* were the case. When a person regularly behaves in these ways, this can provide ground for attributing to him the belief that *p*, even if he is unaware of having such a belief, and even if he would deny having such a belief if he were asked. Similarly, I think that a person can be said to have an *akratic* belief that *p* if he regularly behaves in ways that are best understood as relying on the premise that *p*, even though he judges there to be insufficient

1. Donald Davidson, "How Is Weakness of Will Possible?" in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980), pp. 21–42.

evidence for p and perhaps compelling evidence for not- p . (Think here of a person who acknowledges that there is every reason to think that his son is dead but nonetheless operates on the premise that the son is still alive, making preparations for his return and continuing to set aside money for his college education.) It was this kind of case that I meant to describe in the passage Wallace cites. When I said that someone might, in the presence of an artful deceiver, "find himself thinking" that that person was reliable after all, I meant not only that he was tempted to think this, but that he found himself engaged in reasoning based on the assumption that this was so.

Responsibility

As Wallace observes, I take a broader view than he does of the range of states that are subject to strong moral criticism. Central to this difference is a difference in our views of responsibility (or attributability). In my view, all of the states that are in principle responsive to a person's judgment about reasons (all of his or her judgment-sensitive attitudes) are attributable to that person in the sense required for moral assessment. Wallace, on the other hand, holds that people are responsible only for those attitudes that are in fact under their direct rational control.

Wallace agrees that even wayward desires that are not responsive to an agent's judgment "must in some sense be attributable to the agent, otherwise there would be no special need for the agent to struggle against them" (p. 441). The question is, in what sense are they attributable? I maintain that it is the sense in which all of a person's judgment-sensitive attitudes are attributable to him or her. There is a need to "struggle against" wayward desires just because they involve judgment-sensitive attitudes, that is to say, ones that are in principle responsive to one's judgment. Even the idea that we are "alienated from" such desires presupposes this: we are alienated from them because they conflict with, but resist, our judgment. Without this element of conflict they would be "alien" only in the sense that our height is—that is to say, merely in being beyond our rational control.

Not only moral criticism, but also criticism on personal and intellectual grounds (criticism of our views about chess strategy, for example) is addressed to attitudes that are attributable to us in this same sense. For this reason, all such criticism has a special kind of significance for the person criticized. Unlike criticism of our height or the shape of our ears, criticism of our judgment-sensitive attitudes challenges us to make a decision: to reconsider the judgments in question and to either defend them or modify our attitudes accordingly. Criticism of this kind may be appropriate—the modification of my attitudes that it demands may be called for—even though I do not, perhaps even cannot, see that this is

the case. I am open to criticism for my stupid chess moves whether or not I am able to see that they are stupid.

Wallace might respond (see p. 442) that it may nonetheless not be appropriate to blame me for these moves if my incorrigible strategic blindness prevents me from doing any better, and that, similarly, a person is not to be blamed for his or her faulty moral attitudes if he or she is unable to see that they are faulty, or unable to control them. I agree that it sounds odd to speak of blaming an unskilled chess player for his bad strategic judgments. But this is not because those judgments, being beyond his control, are not attributable to him in the required sense. The reason, rather, is that 'blame' denotes a specifically moral form of criticism, criticism for faults that have a kind of significance for our relations with a person that strategic blunders in chess, in themselves, do not have. Of course, it might be claimed that just because moral criticism has this kind of significance it is appropriate only for things that are under an agent's control. I assume that this is what Wallace would claim, so I need to explain why I disagree.

First, let me say that I take the question at issue to be the conditions under which moral criticism is correct and appropriate, not the quite different question of when it is appropriate to address these criticisms to the person in question by, for example, berating him for his faults or demanding explanations or apologies. Blameworthiness in this sense is the correct thing to focus on in a dispute about conditions of attributability, since if an attitude is not attributable to a person in the required sense this should undermine the judgment of blameworthiness itself rather than merely make it inappropriate to address that judgment to the person.

In my view, a person is open to moral criticism if he does in fact hold judgment-sensitive attitudes that are morally objectionable, such as the attitude that he has no reason to take any account of other people's interests or the possibility of their being harmed by his actions. Such attitudes have the kind of significance for the person's relations with others that is central to moral criticism because, as Wallace puts it well in stating my view, they betray "a kind of disregard, a failure to recognize and respond appropriately to [other people's] value as persons" (p. 442). This is so whether or not the person is capable of modifying these attitudes, or even capable of seeing any reason to modify them. Whether these conditions are fulfilled or not, people have reason to object to these attitudes and to take them as significant for their relations with a person who has them. In my view, to say that a person is blameworthy is just to say that he holds attitudes that are open to moral criticism and have this kind of significance. Wallace's conviction that people are subject to "strong moral appraisal and blame" only for attitudes that are under their rational control may reflect the fact that

he understands blame differently, perhaps as involving an idea that the person deserves to be punished, or to suffer in some other way.

Wallace points out that in my view moral criticism of an attitude calls upon a person to withdraw or modify that attitude, and he argues that this is inappropriate where rational control, or moral capacity, is lacking (pp. 442–45). But this is a slight misconstrual of my position. I do say that moral criticism and, more generally, criticism of judgment-sensitive attitudes, calls upon an agent to reassess his or her attitudes and to defend or revise them. I emphasized this in order to bring out the way in which this kind of criticism differs, from the point of view of the person criticized, from criticism of physical traits such as height and beauty. As I pointed out above, however, criticism of judgment-sensitive attitudes “calls for” this kind of response whether or not the person criticized is capable of making it, or of seeing why, in a particular case, revision is called for.

It is true that from the point of view of a person who was injured by the agent’s action and is criticizing it on this account, *one* of the things that may be desired in *making* that criticism is for the agent to acknowledge his complaint as valid and modify his attitudes accordingly. But both of the italicized words in this sentence are important. First, while this is one of the things that may be sought in making the criticism, it is not the only thing: the injured person may also seek acknowledgment of the legitimacy of her claim from third parties, and she may think it is important simply to affirm her dignity and standing by making it. Second, as I have emphasized above, our reasons for taking moral criticism seriously, and for attaching importance to its validity, go beyond our reasons for expressing it. The attitudes a person holds toward us are significant for our continued relation with him whether or not we have the opportunity to complain to him about them, and whether or not it would be wise, or effective, to do so.

I have argued that “wayward” desires—ones that are in conflict with a person’s considered judgment—can be nonetheless attributable to the person for purposes of moral assessment. But it does not follow that the waywardness of a desire—the fact that the person rejects it and would eliminate it if he or she could—is morally irrelevant and that a wayward desire has the same significance in moral assessment as a similar desire that is fully endorsed. The fact that a desire is in conflict with a person’s considered judgment changes things, not by making that desire *not* attributable to the person, but by providing a more complete picture of what *is* attributable to him or her. It is one thing to feel indifference or loathing toward a person while believing firmly that these attitudes are objectionable, and wishing to be rid of them, and quite a different thing to have the same attitudes with no conflicting judgment. The former is a less serious fault, perhaps in some cases only a very minor

one. Some might say that if the person has, without success, done everything possible to rid herself of this wayward attitude, continuing to have that attitude is no fault at all (that criticism for it is inappropriate). But this is most coherently understood as a substantive view about the standards of perfection that it is reasonable to hold people to, rather than as the claim that the wayward attitude is not attributable to the person in the relevant sense. The latter position is difficult to square with the fact that the person herself would reasonably see this attitude as a fault that she would like to be rid of. Moreover, the spontaneous responses of friends and lovers are at least as significant for us as the attitudes that survive reflective self-correction. How is this to be accounted for if these wayward spontaneous responses do not belong to the person?²

Value

I agree with Wallace's reinterpretation of my "buck-passing" account of value. My thesis was that goodness is not itself a property that provides reasons, not that the underlying properties that do this are always natural properties, and I should not have written in a way that suggested this. He is quite right that more specific evaluative properties often play this role.

The Priority of Moral Reasons

I am grateful to Wallace for his perceptive reconstruction of my response to the problem of priority. He is certainly correct that the idea of justifiability to others, viewed in a certain way, can seem rather cold and abstract. As he concludes, my intention was to portray this idea in a warmer and more robust light, as a way of living which is appealing and valuable. No doubt I should have taken care to be more consistent in this portrayal.

Although I hoped to describe the morality of right and wrong in a way that made clear why commitment to it is valuable, and although I recognized that an explanation of the priority of morality over other values would sometimes come down to what he calls my second strategy (head-to-head comparison of the importance of the competing values), my aim was to minimize the range of cases in which it would come to this. I hoped that in most cases the priority of moral reasons could be explained by the combination of what he calls my first strategy (showing

2. I discuss these matters at greater length in "Reasons and Passions," in *Contours of Agency: Essays for Harry Frankfurt*, ed. Sarah Buss and Lee Overton (Cambridge, Mass.: MIT Press, 2002), pp. 165–83. For a thorough treatment of the question, see Angela Smith, "Attitudes, Agency, and Responsibility" (Ph.D. diss., Harvard University, 1999). I am much indebted to her for discussion of these issues.

how contractualist morality makes room for projects and commitments) and my third strategy (showing how other values have a built-in sensitivity to moral requirements).

I would add one thing to what Wallace says about my discussion of friendship. As he says, I defend a "moralized" version of friendship which, because it involves recognizing a friend as a person with moral standing independent of the friendship itself, is implicitly committed to the moral standing of nonfriends. I believe that this is the best interpretation of friendship as we understand it and that it is the form of friendship most worthy of adoption (because of the way it portrays friends themselves, not because it is less threatening to others).³

I do not claim that this is the only thing to which the name 'friendship' could properly be attached. This admission does not seem to me to weaken my argument as much as Wallace suggests, although he is right that I should have been clearer about what I was claiming. One thing I said was that if my claims about the self-limiting character of this particular form of friendship are correct, then there is at least one form of friendship worth valuing that contractualist morality does not rule out. I think this is a point worth making, but I realized in reading Wallace's comments that it responds to a challenge that differs from the problem of priority. If successful, this argument rebuts the charge that a commitment to the priority of moral requirements would impoverish our lives by depriving us of the possibility of having friendships worthy of the name. But it does not address the problem faced by a person who has a different ("unmoralized") notion of friendship and finds that it conflicts with his moral obligations to strangers. Here I think I have to fall back on a combination of the second strategy with the third, and argue (1) that the form of friendship we have most reason to value is this "moralized" version, which does not make demands that are inconsistent with contractualist morality, and (2) that other forms of friendship, which do make such demands, are for that reason less worthy of being valued and ought therefore to be sacrificed when conflicts with morality arise.

The Unity of Morality

One of my hopes in writing this book was that it would lead to wider discussion of the question of the unity of morality, and I am very pleased that Wallace takes up this issue. Both within moral philosophy and in wider public discussion, people commonly speak of "morality" as if it were one unified normative domain, and I believe that more discussion

3. See *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), p. 165.

of the degree to which this is in fact the case would be a good thing both for moral theory and for political discourse as well.

I do not see any tension between my contractualist account of right and wrong and my more general account of how we should think about reasons and value. As Wallace observes, the former has a definite structure that the latter lacks. But this structure is a feature of the particular value that is being examined and articulated. The general method that I describe in chapters 1 and 2, when applied to values such as friendship, or various forms of aesthetic value, might well reveal that these values too have their own distinctive structures, as revealed, for example, in criteria of relevance and irrelevance, or requirements of generality, or the lack of such requirements.

Wallace is correct that there is a certain instability in my position on the question of unity. Once I have argued that morality, as it is commonly understood, involves many diverse values, the question arises whether a similar kind of reflection will not show that what I call the morality of right and wrong is best understood in terms of diverse moral values and thus lacks the unity that I claim for it. Although I do not think this is the case, I do see the challenge. But I understand it in a somewhat different way than Wallace appears to. To explain how, let me consider three ways in which more specific values might play a crucial role in our thinking about right and wrong.

The first such role is as something that individuals have reason to regard as important and to appeal to as grounds for rejecting proposed principles of conduct. As Wallace mentions, the value of assurance plays this role in my argument for the principle of fidelity. I would not, however, call assurance a "thick ethical concept." It is simply something that people in certain circumstances have reason to want: the ability to rely on another person's performance. Fairness, on the other hand, may be an ethical concept, but I rely on it in the same way as assurance: as something that people have reason to want, or to object to the lack of. I argue that people have reason to object to being deprived of benefits arbitrarily—that is to say, without justification—and that this constitutes a good reason to reject principles that would leave them open to such treatment.

The second possibility is that an ethical concept may be a way of expressing a particular set of conclusions about what is right and wrong in the way contractualism describes—conclusions about what would be allowed or disallowed by nonrejectable principles. Fidelity, for example, may be understood not as a value or virtue independent of the idea of justifiability to others but as a consequence of this idea: it is the kind of conduct that is required by principles that describe a part of "what we owe to each other." The same might be said of murder, understood as unjustifiable intentional killing.

The third role is that of a free-standing ethical value or ideal. By a "free-standing" ideal, I mean one that is independent of the idea of what we owe to each other that contractualism describes. So, for example, a person who has a certain conception of honor may keep his word not because he owes this to the person to whom he gave it, but because it would be dishonorable, or weak, to fail to do so. Ideals of this kind may also serve as reasons for rejecting principles, in the way described above. I might, for example, object to a principle because it did not give me the freedom to live up to my ideals. Whether this would be reasonable would, of course, depend on how others would be affected by my having this freedom. But what is important for the moment is not this possible role within contractualist argument but, rather, the fact that these ethical ideals may represent components of morality independent of the value that contractualism describes.

Some of Wallace's examples seem to me to be of the first two kinds, but only the third possibility seems to me to present a challenge to the unity of the part of morality that I am describing. I do not deny that there are free-standing ethical concepts of this sort. The question is how central they are to our thinking about right and wrong. I believe that once the independence of such ideals from the idea of justifiability to others is made clear—as, for example, when the idea of honor is clearly distinguished from fidelity—these ideals cease to seem central to our moral thinking (even though they may not lose their appeal altogether). But this is a matter that has to be decided by carrying out the analysis in each case, and it depends on how convincing one finds the contractualist account of concepts such as fidelity.

REPLY TO DWORKIN

I am grateful to Dworkin for undertaking to respond to Philip Pettit's criticism of my view, and I agree with much of what he says in this response. There are, however, a few things I would like to add in the hope of clarifying my position.

The problem of explaining moral motivation, or, as I would prefer to put it, explaining the normative basis of right and wrong, is naturally seen as a matter of explaining how the fact that a certain action would be wrong provides a reason not to do it. I myself sometimes put the matter this way, but as I point out in my chapter 4, this is somewhat misleading.⁴ The normative basis of morality is not simply a special reason, analogous to a legal sanction, that is triggered by the conclusion that an action would be wrong and then (at least normally) outweighs competing considerations. A good person is moved to do what morality requires not only "because it would be wrong not to" but also by more

4. See *ibid.*, pp. 147 and 155–58.

concrete considerations such as the fact that acting in a certain way would kill or injure someone, or the fact that someone needs help, or the fact that someone is counting on him or her to do something. Particular right-making and wrong-making considerations such as these will be uppermost in the mind of a morally good agent on many occasions, with thoughts such as "it would be wrong not to" serving mainly as normative backstops in cases of temptation.

But particular right-making and wrong-making considerations such as those I have just listed are numerous and diverse, and can support conflicting directives. So acting morally cannot be identified simply with recognizing them as important and being moved by them. We need some way of deciding when they are relevant and when they are decisive. Moral thinking involves answering such questions as, What measures must I take to avoid possible injury to others? Does this inconvenience count as a good reason to disappoint someone's expectation? Is it permissible to impose this cost on one person in order to benefit others in that way? In addition to the "backstop" function mentioned above, moral motivation has an important role in moving us to ask questions such as these and in shaping the way in which we go about answering them.

A moral theory is a systematic attempt to describe the process through which we should go about answering such first-order moral questions and to explain why we should regard this process and its results as having special importance. I will refer to these tasks as, respectively, the task of explaining the content of morality and the task of explaining its normativity. The issues at stake in Pettit's criticism of my view can be put in terms of several contrasting sets of answers to these two theoretical questions.

I hold that the moral permissibility of an action depends on whether a principle that permitted it would, for that reason, be one that someone could reasonably reject. In my view, we have reason to be concerned with the moral permissibility of our actions because we have reason to be concerned with the justifiability of our actions to others. Pettit holds, by contrast, that to determine what morality requires we must ask what will lead to maximum value, and that we have reason to be concerned with morality because we have reason to promote maximum value.

I disagree with both parts of his view. It does not seem to me that right and wrong can be adequately described in terms of the maximization of value. But Pettit's criticism of my view does not merely repeat his side of this disagreement. He says that my view will be deeply implausible unless the question of content can be answered by giving a general characterization, in substantive terms, of the way in which considerations such as death, injury, frustrated expectation, and so on, are to be taken into account and reconciled with one another in order to yield correct answers to questions of right and wrong. His criticism does

not assume that this characterization must be put in terms of maximizing value but only that it has to be specified in substantive terms (i.e., in terms of the underlying values of preserving life, avoiding injury, etc.) and specified in a way that is independent of and prior to ideas such as reasonable rejection or justifiability to others. Once such a characterization has been given, it might then be said that it would be reasonable to reject a principle if that principle did not reconcile these underlying values in the correct way.

Dworkin responds that I do not need to deny the possibility of a substantive characterization of right and wrong of this kind. Even if there were such a characterization, there would still be some work for the contractualist idea of justifiability to others to do in answering the question of normativity. I could say that we have reason to take this substantive criterion seriously as a guide to action because actions that are not in accord with it are ones that could not be justified to others.

I do not reject the possibility of such a hybrid view.⁵ But I doubt whether such a substantive characterization of right and wrong is in fact available. I do not see any general way of specifying in substantive terms, across the range of cases included within the sphere of morality that I call what we owe to each other, exactly which specific considerations are relevant in which cases and how these are to be reconciled with one another. Although we can arrive at many conclusions about which principles dealing with specific classes of cases it is or is not in fact reasonable to reject—principles having to do with promises and expectations, principles having to do with safeguarding life, and so on—I think there may be no way of giving a substantive answer to such questions that applies generally, across the whole range of cases. In particular, familiar objections to consequentialism seem to me to show that the idea of maximizing the value of resulting states of affairs does not provide such a general account. The idea of reasonable rejection may be as good a general guide as we can get. I do not see why my view should be so deeply implausible just because this turns out to be the case. Of course, someone who believes that there is such a general substantive account (e.g., one stated in terms of maximizing value) may believe that taken by itself it provides a fully satisfactory moral theory. Such a person would presumably reject contractualism. But this would be a different objection than the one Dworkin is discussing, since it would depend on the case for accepting that particular substantive view.

As Dworkin notes, my answer to the question of content refers to principles that no one could reasonably reject, *insofar as they were moved*

5. I discussed this possibility on p. 118 of "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, ed. Amartya Kumar Sen and Bernard A. O. Williams (Cambridge: Cambridge University Press, 1982).

to find principles that others who share this aim also could not reasonably reject. As I make clear in my book, I believe that everyone has reason to seek and be guided by such principles. One may wonder, then, why I include the italicized motivational condition. Including it might have a point if I were leaving open the possibility, as I did in "Contractualism and Utilitarianism," that people who lack a desire to act in ways they can justify to others may lack a reason to do so. This possibility is not left open in *What We Owe to Each Other*; so the motivational condition may appear to be a residue from that earlier work—something that once served a purpose in my view but is now redundant.

It is not redundant, however, for a reason that has to do with the idea of reasonableness that I am employing. A claim about what it is reasonable to do, or believe, presupposes a certain contextually specified body of information and a certain range of reasons, and makes a claim about what these considerations taken together in fact support. So, for example, a claim about what it is reasonable for someone in a certain situation to believe is a claim about what is supported by the evidence available to someone in that situation (not necessarily all of the evidence that is in fact relevant to the question). The function of the motivational condition mentioned above is exactly this kind of contextual specification: it is to indicate that the aim of finding principles that others, similarly motivated, could also not reasonably reject is one of the considerations that are to be taken into account in assessing the reasonableness of rejecting a principle. (Other considerations are specified by the generic reasons that a person is taken to have in virtue of occupying a certain position in situations of the kind to which the principle under discussion is to apply).

The content of the morality of right and wrong is determined, in my view, by considering what principles it would or would not be reasonable to reject from various standpoints defined in this way. I address the related question of normativity in two ways. First, I claim that we have reason to care about the justifiability of our actions to others, and one aim of my description of the content of morality is to make clear how this reason is central to its normative basis. Of course, it is difficult to argue for the claim that we have such a reason, other than by describing it in a way that one hopes will make clear its appeal. But even if it is admitted that we do have such a reason, it is not the only reason we have. So the second part of my response to the question of normativity consists in showing how the justifiability of our actions to others is related to other values, with the aim of showing how it can make sense to give the requirements of justifiability priority over these other concerns.⁶

6. I undertake this in sec. 5 of chap. 4, *What We Owe to Each Other*, pp. 165–68.

Dworkin raises the question whether "the disaffected"—those who are not concerned with the justifiability of their actions to others and deny that they have any reason to be so concerned—nonetheless have such a reason. This is in my view a substantive question about reasons. Dworkin is correct that I would here invoke what I call the universality of reason judgments. This is the thesis that other people's reason to X is dependent on some condition C being satisfied in their case if and only if our having that same reason depends on C being satisfied in our own case. So, in particular, in order to decide whether "the disaffected" have reason to be concerned with the justifiability of their actions to others it suffices to decide whether our own reason for being concerned with the justifiability of our actions depends on our in fact caring about this. Such a question can be answered only by examining the reasons in question and making a judgment about what conditions they depend on.

Dworkin appears to agree that our reasons for not killing others, and for not being cruel to them, do not depend on our happening to care about these things. But he says that such examples are "at the wrong level." What he means by this, if I understand him correctly, is that the claim that we have motivation-independent reasons to avoid killing and cruelty may be internal to our moral view, and that we may have such reasons just because we accept that overall view. So the fact that these reasons are independent of our motivation does not show that our reasons for accepting this overall moral view are similarly independent.

But it seems to me just as clear that our reasons for being concerned with morality at all (for caring about the justifiability of our actions to others) are not conditional on our happening to care about this and to be moved by it. These reasons are not themselves moral reasons (that is to say, it is not that we have reason to care about the justifiability of our actions because it would be morally wrong, or morally bad, not to do so).

The general claim that something is a reason for a person only if it is, or follows from, something he or she is moved by is not usually defended as a substantive claim. Arguments for this thesis tend to appeal instead to what are alleged to be more formal constraints on reasons, such as the idea that if C is a reason for A to X, then this must be something that could be offered to A as advice, and that it is inappropriate (mere browbeating) to offer as advice something that has no basis in that person's motivations. I counter this argument by pointing out that from the fact that it would be inappropriate to say something to a person (e.g., inappropriate to go on repeating an argument that he has already rejected), it does not follow that what you would say if you were to do this would not be true (or good advice). Moreover, if it is a deficiency on a person's part not to be able to see that C is a reason for him to X, this seems to imply that it is true that it is a reason. Cruelty, for example, involves a failure to see other people's pain as something one has a reason

to avoid. If cruelty is a deficiency, this must be because the pain of others is in fact a reason—a consideration that counts against the cruel person's acting in certain ways—even though he or she (being cruel) fails to be moved by it.⁷

At the end of my appendix on Williams, I consider that a defender of the view that reasons depend on motivation might accept this point about deficiency, agreeing that there is a reason for even a cruel person to avoid causing pain to others, and that, more generally, whether some consideration C is a reason for A to do X does not always depend on A's being moved by C. The area of disagreement between this position and one like mine would then be narrowed significantly. We might agree on the question of when C is a reason for A to X and disagree only over whether the phrase "A *has* a reason to X" should be restricted to cases in which the reason is one that is available to A—one that A is capable of seeing and being moved by. As I make clear in that appendix, I do not myself favor this restriction, but I suggest that this residual disagreement would not be a very important one. Dworkin disagrees, saying that concession on this point would undermine moral criticism and make moral reactive attitudes inappropriate.

In saying that this disagreement about the range of "has a reason" would not be very important, I was assuming that the applicability of moral standards to a person would depend on the broader idea of there being good reason for him or her to accept these standards rather than the narrower idea of this reason being one that that person was capable of appreciating. I believe that Dworkin does not accept this assumption, for reasons that bring us back to issues I have already discussed in my reply to Wallace. I will not repeat that discussion here but will only add one point about reactive attitudes.

These attitudes are called reactive because they are reactions to the attitudes toward us that are manifested in someone's actions. Resentment, for example, is a reaction to actions of others that manifests a lack of concern, or even contempt, for our interests. A reactive attitude is appropriate, I would say, just in case the person's behavior really does manifest attitudes toward us of the relevant kind. Various factors can qualify or undermine the degree to which this is the case. For example, although the terseness of someone's response to our problems may seem to indicate a lack of concern and sympathy, the fact that the person was very tired or under great strain may modify the degree to which these attitudes are properly attributed to him. But the fact that a person is

7. Williams discusses this point and offers a somewhat different interpretation of the sense in which cruelty is a "deficiency" in a postscript to "Internal and External Reasons," in *Varieties of Practical Reasoning*, ed. Elijah Millgram (Cambridge, Mass.: MIT Press, 2001), pp. 95–96.

unable to see why our interests should matter to him does not similarly undermine the degree to which his actions can be said to manifest the view that our interests do not matter. So I do not see that the reactive attitudes would be rendered inappropriate by the restriction I described on the range of cases in which a person can be said to have a reason.

REPLY TO DEIGH

John Deigh's objections to my account of promissory obligation are of two kinds. First, he presents a number of counterexamples, mainly cases arising in the context of war that he believes my view cannot explain. Second, he uses these examples to raise a version of the theoretical objection that in the absence of a convention of promise making there would be no way to induce the kind of reliance that is required, on my view, to generate an obligation to keep a promise. I will reply to each of these in turn.

As Deigh notes, the principle that I offer to account for the obligation to keep a promise (principle F) is the last member of a sequence of principles concerning the expectations we lead others to form about what we will do. The earlier members of this sequence deal with obligations not to intentionally mislead others about our intentions (principle M), to take due care not to lead others to form false expectations about what we are going to do (principle D), and to take reasonable steps to protect others against losses that they might suffer as a result of relying on false expectations about what we are going to do that we have intentionally or negligently led them to form (principle L).

The sources of these three principles (the reasons why it would not be reasonable to reject them) lie in the reasons people have to want to avoid suffering losses as a result of relying on false expectations. The obligation to keep a promise cannot be accounted for solely on this basis, however, since one can be obligated to keep a promise even when the person to whom it was made has not yet made any decision in reliance on it. Accounting for obligations of this more extensive kind requires a further principle, which requires a different reason to support it. (For this reason, I did not claim that principle F "falls out of" principle L, as Deigh suggests.)

In addition to having reason to avoid reliance losses, we also often have reason to want to be sure that others will behave in a certain way (unless we consent to their not doing so), whether or not we have made any choices relying on the assumption that they will do this. Suppose, for example, that there is only one company from which I can buy hurricane insurance, and that there is nothing I can do to minimize storm damage to my house. In these circumstances I would have reason to pay a higher premium in order to get a policy that the company could not cancel at will. This policy would provide what I called the

value of assurance. I appealed to this value in arguing that it would not be reasonable to reject principle F, which applies in situations in which certain conditions of mutual knowledge and expectation are fulfilled, and requires one to do what one has, under these conditions, led someone to expect that one will do.

I believe that the conditions of principle F are fulfilled in most cases in which promises are offered and accepted, and that this principle therefore explains the obligation to keep a promise. The alternative explanation that Deigh favors holds that this obligation is based on the obligation to abide by the rules of a social practice of promising. The explanation that principle F provides seems to me more plausible, in part because it accounts for the ways in which the wrong of breaking a promise is importantly similar to, and also different from, the wrongs involved in other cases in which someone intentionally creates false expectation in others, or disregards expectations he or she has created. These wrongs (violations of other principles in the family I have described) need not involve social practices. Yet they bear a deep moral similarity to the wrong involved in making a false promise or failing to keep a promise made in good faith.

To maintain that the obligation to keep a promise does not depend on a social practice of promising is not to deny that there is such a practice, much less that there are many social practices of other kinds. Where there are such practices, there can also be obligations to comply with the rules that define them, obligations that can be explained on, for example, Rawlsian, Humean, or rule utilitarian grounds. In my book, I affirm that there is a social practice of promising at least to this extent: there is a linguistic practice according to which sincere use of the words "I promise . . ." indicates a belief that one is in a situation of the kind described in principle F and an intention to offer assurance of the kind that principle describes. Insofar as that principle is accepted as a norm (even if only for the reasons I offer) there will also be a social practice of promising in a slightly stronger sense, consisting of the acceptance of this norm, according to which one is normally required to do what one has promised.

If the acceptance of this norm consists merely in the general recognition of the validity of a moral argument of the kind I offer, then there is nothing in this that is in any way in competition with my view. The generally accepted norms of agreement making might, however, not be seen in this way as independently valid moral requirements but, rather, as rules that are morally obligatory in virtue of their general acceptance and social usefulness. The existence and moral bindingness of such a practice would be quite compatible with my account of promissory obligation. It could well be true that a person who promises to do something commonly incurs two kinds of obligation to do it: one of

these an obligation to abide by and support a useful social practice, the other an obligation of the kind principle F describes. (The first of these obligations would in fact support the second, since the practice would provide a way of creating expectations of the kind that are central to principle F.)

One way in which the difference between these two forms of obligation to keep a promise would be shown is in their similarities to other obligations. The second obligation belongs to the same family of obligations as those described by principles M, D, and L, all having to do with the expectations one creates, while the first has (at least on the Rawlsian view) the same source as the obligation not to free ride on useful institutions or to undermine them by spreading rumors about other people's noncompliance.

A second difference between the two obligations lies in the ways that questions about their extent (e.g., about the validity of exceptions to them) are to be settled. In one case, such questions will be a matter of basic moral argument; in the other case, they will at least sometimes be questions about what our social practice in fact is (given that it might, morally speaking, have taken any one of several forms). The apparent failure of our promissory obligations to depend in this way on morally arbitrary social facts about "the rules of our practice" was one thing that inclined me to doubt that a social practice has a stronger role of this kind in generating our promissory obligations.⁸ Even if one held that there is a practice that plays this role, however, this would not be incompatible with my account of promises. All that I deny is that such an institution is essential. I will return to this point below.

Let me turn now to what Deigh says about promises between combatants in time of war. First, I note that principles M, D, L, and F do not require the same thing under all conditions. In particular, principles M and F forbid certain conduct only "in the absence of special justification."⁹ Just after using this phrase for the first time, in the statement of principle M, I explain what I mean by it, citing four types of cases. Three of these cases are relevant here. These are, respectively, "emergency cases," in which it is necessary to deceive someone in order to prevent something very bad from happening to others; "threat cases," in which B presents an illegitimate threat to someone (the example I give is kidnapping) who can only be saved by misleading B; and "permission cases," in which the

8. See *What We Owe to Each Other*, pp. 309–10. The other main reason was, as I have said, the apparent moral similarity between the wrongs involved in insincere or broken promises and those involved in violations of principles M, D, and L.

9. I probably should have included this in principles D and L as well, but did not think this was necessary given that they are already qualified by requiring only "due care" and "reasonable steps" rather than more specific actions.

parties have entered, by mutual consent, into a game or other activity, such as a special form of competition, that is understood to be one in which certain forms of deception are permitted.¹⁰

As Deigh says, enemies in war have no general obligations to forbear from deceiving and manipulating each other in ways that principles M, D, L, and F would generally prohibit. Duties to refrain from killing and injuring others are also normally taken to be suspended in wartime, at least between combatants. The claim that they are is not uncontroversial, but it is most plausibly justified, I would say, along one or more of the lines I have just described: violence and deception are justified in war either by the need for self-defense against a threat or by the idea that war is special kind of interaction governed by a convention according to which certain (but perhaps not all) moral restraints are suspended. I suspect that explanations of both of these kinds are required, in different cases.

It would be foolish to deny that there are conventions governing conduct during war, and nothing I say in my book suggests this. (Since these conventions are explicitly codified, the ground I mentioned above for doubts about the role of conventions in regard to practices—that they have no identifiable content beyond that of ordinary moral principles—does not apply.) As I have just suggested, however, these conventions need to be understood as licensing departures from ordinary moral restraints as well as imposing special restrictions. Indeed, it is only against a background in which violence is specially licensed that the two restrictions Deigh mentions—the use of white flags and the Red Cross insignia—can be properly understood. They represent conventionally established limitations on the conventionally licensed use of violence between combatants in war. Showing a white flag is not merely a way of inducing one's enemy to have a certain expectation (although it may do that). It also invokes a convention making it wrong for them to attack you, or to refuse to give you the status of prisoner of war.

These things could clearly not be explained on a basis that did not take the conventions of war into account, and nothing I say in my book suggests that they should be. Perhaps Deigh's claim that I "favor" a different account is based on the fact that where these conventions are generally accepted, symbols such as white flags and red crosses will create expectations and can be used to mislead and manipulate. He may conclude from this that I am committed to the view that misuse of these symbols in wartime is wrong simply because it violates my principle M. This would be a mistake: as I have explained, M does not have this implication. But Deigh is correct that the result would be implausible. For example, if using the Red Cross symbol to disguise trucks carrying

10. See *What We Owe to Each Other*, p. 299.

ammunition were impermissible because it violates M, then sending out decoy trucks to tempt the enemy to attack in a place disadvantageous to them would also be wrong for the same reason. Since this is clearly not wrong, the wrongfulness of misusing the red cross must, as Deigh says, be explained in some other way. The most plausible explanation, as he suggests, is not simply that it misleads but that it inappropriately invokes a conventional prohibition, in a way that cannot be permitted if that convention is to do its useful work. The same is true, I would say, of showing a white flag in feigned surrender in order to get enemy soldiers into an exposed position where they can be attacked.

The general point here is this. Deigh seems to suppose that because I deny that a social practice of agreement making is essential to the kind of obligation that a promise can create, I must deny that conventions ever play a role in explaining our obligations. My reply is that I do not deny the importance of conventions. They are important not only in creating expectations (as in the account of promises that Deigh favors) but also in selectively licensing forms of behavior that are normally prohibited. It seems to me that Deigh does not give sufficient attention to the latter function of conventions and, in particular, to its selective character.

This selectivity is an element in one of the questions raised by agreements entered into by defeated nations. One possible problem about such agreements is that they may seem to be invalid because they are entered into involuntarily, that is to say, only because of the absence of acceptable alternatives. As Hume noted in a passage I cite in my book, the same might be said about a promise to pay a surgeon.¹¹ My response, as Deigh mentions, is that voluntariness, in the sense relevant here, is a moral notion. Whether a choice is involuntary in this sense depends not simply on the nature of the alternatives available to the agent, or the agent's attitudes toward these alternatives, but also on whether the agent has been deprived of alternatives to which he is entitled.¹² The fact that a nation would not have accepted a treaty but

11. David Hume, *Treatise of Human Nature*, bk. 3, pt. 2, sec. 5. See my *What We Owe to Each Other*, p. 245.

12. The term 'voluntarily' is somewhat misleading here, insofar as it suggests a distinction based entirely on the nature of the alternatives and the agent's attitude toward them. For this reason, in later work I have revised principle F so that instead of saying "If A voluntarily and intentionally leads B to . . ." it says "if A, in the absence of objectionable constraint and with adequate understanding (or the ability to acquire such understanding) of his or her situation, intentionally leads B to . . ." ("Promises and Contracts," in *The Theory of Contract Law*, ed. Peter Benson [Cambridge: Cambridge University Press, 2001], p. 95. The idea of voluntariness is discussed further in sec. 5 of that paper, pp. 111–17). Deigh observes that in chap. 6 of my book I describe a broad notion of responsibility according to which a person is responsible for any action that can correctly be attributed to him, and he says that I "plainly return to the ordinary conception in

for the fact of military defeat does not prevent that treaty from being morally binding. Whether it is so or not depends on the terms it involves.

A second problem that Deigh raises concerns whether, in negotiations between enemies at the close of a war, one party would have a reason to believe in the other's good faith that would be sufficient to generate an expectation of the kind required to create an obligation under principle F. As he says, this is an instance of the more general problem raised by Elizabeth Anscombe. Replying to Anscombe's objection in my book, I argue that if, in a situation of the kind F describes, a person purports to be offering me assurance that he will do X, although he in fact has no intention of doing this, then he is misleading me about his intentions in a way that contravenes at least principle D and perhaps M as well. Therefore, if I have reason to believe that I am dealing with a morally serious person, who recognizes these principles and would not violate them, then I have reason to believe that he intends to follow through on his assurance if I accept it. Deigh objects that this argument falls through in situations in which principles M and D do not rule out deception, as I have admitted that they do not between combatants in war. He goes on to mention other cases, such as adversarial bargaining, in which I would also agree that deception is licensed. The question is how the expectations required to generate an agreement that is binding according to principle F could be generated in the way I have just described in situations in which it is known to be permissible to mislead others about one's intentions.

Even if the answer were that this is not possible, this would not entirely undermine my response to Anscombe's objection. It would just show that there are some conditions under which the expectations required to generate a binding agreement can be generated only with the help of "external" (that is to say, nonmoral) factors, such as legal or other mechanisms of enforcement, or the disapproval of third parties. This would still be an unwelcome result, but only a limited one.

Deigh's argument does not, however, establish even this limited result. It appears to do so only because he assumes that the permission to mislead in situations of the kinds he describes would be a blanket permission. As we have seen above, however, it is more plausible to suppose that this permission is selective.¹³ In wartime it is permissible

Chapter 7." This suggestion of inconsistency is misplaced. One of the main points of chap. 6 was to emphasize the distinction between the broad notion of responsibility that Deigh mentions, which I called responsibility as attributability, and a different notion, which I called substantive responsibility. I argue that it is the former notion that is a precondition for moral assessment, and that the latter is the one that is relevant when, for example, we are asking whether an agent can be required to bear the costs of a certain choice. (See chap. 6 in general and esp. pp. 248–67 and 290–91.)

13. In a passage that Deigh quotes, I make just this point about lying (p. 320).

to deceive one's opponents on the battlefield in certain ways. But why suppose that this extends to a permission to deceive them about one's sincerity in entering into a peace agreement? What purpose would this permission serve? Similarly, why think that the permissibility of misleading others about one's "bargaining position" (by saying such things as "this is my final offer") extends also to misleading them about one's willingness to abide by the offers one makes? Again, what purpose would this serve?

Deigh imagines that because I do not believe that obligations of the kind that promises involve require the existence of a social convention I must deny the relevance of social conventions more generally. But I do not deny this. I agree that conventions have a crucial role both in creating and in shaping the obligations we have. But it does not follow from this that all promissory obligations depend on a social practice in the way that Deigh suggests.

Brute Luck, Option Luck, and Equality of Initial Opportunities*

Peter Vallentyne

INTRODUCTION

In the old days, material egalitarians tended to favor equality of outcome advantage, on some suitable conception of advantage (happiness, resources, etc.). Under the influence of Dworkin's seminal articles on equality,¹ contemporary material egalitarians have tended to favor equality of brute luck advantage—on the grounds that this permits people to be held appropriately accountable for the benefits and burdens of their choices. I shall argue, however, that a plausible conception of egalitarian justice requires neither that brute luck advantage always be equalized nor that people always bear the full cost of their voluntary choices. Instead, justice requires that initial opportunities for advantage be equalized—roughly along the lines suggested by Arneson and Cohen.² Brute luck egalitarianism and initial opportunity egalitarianism are fairly similar in motivation, and as a result they have not been adequately distinguished. Once the two views are more clearly con-

* I'm grateful for the helpful comments from Dick Arneson, Ian Carter, Paula Casal, Jerry Cohen, Roger Crisp, Tony Ellis, Cécile Fabre, Marc Fleurbaey, Klemens Kappel, Kasper Lippert-Rasmussen, Andrew Mason, Colin Macleod, David Miller, Mike Otsuka, Ingmar Persson, Terry Price, Eric Rakowski, Arthur Ripstein, Martin Sandbu, Seana Shiffrin, Hillel Steiner, Bertil Tungodden, Alex Voorhoeve, Andrew Williams, and an anonymous editor of this journal.

1. Ronald Dworkin, "What Is Equality? Part 1: Equality of Welfare" and "What Is Equality? Part 2: Equality of Resources," *Philosophy & Public Affairs* 10 (1981): 185–245, 283–345.

2. Richard Arneson, "Equality and Equal Opportunity for Welfare," *Philosophical Studies* 56 (1989): 77–93, and "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare," *Philosophy & Public Affairs* 19 (1990): 158–94; G. A. Cohen, "On the Currency of Egalitarian Justice," *Ethics* 99 (1989): 906–44.

Ethics 112 (April 2002): 529–557

© 2002 by The University of Chicago. All rights reserved. 0014-1704/2002/11203-0006\$10.00

trasted, equality of opportunity for advantage will, I claim, be seen to be a more plausible conception of equality.

BACKGROUND

The concept of justice is construed in several different ways: as rightness of institutions, as rightness of distributions, as coercively enforceable duties, as what we owe others (as opposed to what we owe ourselves or owe impersonally), and as fairness (i.e., what we owe others in purely comparative terms). Here I am concerned with justice in the sense of what we owe others. So understood, justice is not concerned exclusively with the comparative issue of ensuring that individuals with equal claims get equal benefits (fairness). It is also concerned with ensuring that the claims of individuals are fully met (noncomparative justice).

A plausible conception of justice requires, I shall assume, some sort of promotion of equality of material advantage (perhaps constrained by certain rights). Obviously, this assumption is highly controversial. Some (e.g., traditional libertarians) would reject the relevance of consequences generally and consequences for the disadvantaged in particular. Others would defend a sufficiency view, according to which justice requires that everyone have a sufficient level of benefits but does not further require equality.³ Others would defend a prioritarian view, according to which justice requires promoting benefits but with extra weight given to those who are worse off.⁴ Still others would defend some kind of equality of social status quite independently of the effects on material advantage.⁵ My goal here, however, is not to defend the relevance of material equality but, rather, to defend a particular conception of material equality as the most plausible for the theory of justice.

There has been a great debate about the appropriate equalisandum for material equality (and more generally about the appropriate object

3. See, e.g., Harry Frankfurt, "Equality as a Moral Idea," *Ethics* 98 (1987): 21–43.

4. See, e.g., Derek Parfit, "Equality or Priority," Lindley Lecture (University of Kansas, Department of Philosophy, 1991); Dennis McKerlie, "Egalitarianism," *Dialogue* 23 (1984): 223–38, and "Equality or Priority," *Utilitas* 6 (1994): 25–42. Arneson tentatively endorses prioritarianism over (his formerly endorsed) initial opportunity egalitarianism in his "Equality," in *A Companion to Contemporary Political Philosophy*, ed. Robert E. Goodin and Philip Pettit (Oxford: Blackwell, 1993), pp. 489–507, and in his "Equality of Opportunity for Welfare Defended and Recanted," *Journal of Political Philosophy* 7 (1999): 488–97. Arneson's endorsement of (desert-weighted) prioritarianism is explicit in his "Luck Egalitarianism and Prioritarianism," *Ethics* 110 (2000): 339–49.

5. For example, David Miller, "What Kind of Equality Should the Left Pursue?" in *Equality*, ed. Jane Franklin (London: Institute for Public Policy Research, 1997), pp. 83–99; Elizabeth Anderson, "What Is the Point of Equality?" *Ethics* 109 (1999): 287–337; Jonathan Wolff, "Fairness, Respect, and the Egalitarian Ethos," *Philosophy & Public Affairs* 27 (1998): 97–122; and Andrew Mason, "Equality, Personal Responsibility, and Gender Socialisation," *Proceedings of Aristotelian Society* 100 (1999–2000): 227–46.

of concern for justice). Some (e.g., Dworkin, Rakowski, Steiner, and Van Parijs) ground the equalisandum primarily in the competitive value of resources.⁶ Others (e.g., Arneson and to some extent Cohen) ground it in well-being.⁷ Others (e.g., Sen and to some extent Cohen) ground it in functionings.⁸ I shall here leave this issue open and simply appeal to advantage, where this is whatever aspect of outcomes ultimately matters for individuals.

There is, however, a second dimension to this debate. For, no matter what conception of outcome advantage one adopts, there is still the further question of whether it is equality of outcome advantage, equality of opportunity for advantage, equality of brute luck advantage, or something else that is the equalisandum. This article focuses on this second issue.

BRUTE LUCK VERSUS OPTION LUCK

In the next section, I shall argue in favor of equality of initial opportunities for advantage and against equality of brute luck advantage. First, however, I shall examine, in this section, the distinction between brute luck and option luck. For this distinction can be, and has been, drawn in several different ways. It will therefore be useful to start by trying to sort out some of these issues. Throughout, option luck is understood as the complement of brute luck.

Brute luck egalitarianism requires that brute luck advantage be equalized. But what is brute luck? Authors (including myself) sometimes write as if brute luck is to be understood in the following terms:

Brute Luck as Not Foreseeably Chosen: The occurrence of an event is due to brute luck for an agent if and only if the possibility of its occurrence was not (for the agent) a (reasonably) foreseeable outcome of his or her choices.

This understanding of the distinction between brute and option luck, however, is inadequate. To see the problem, compare two cases in each of which there are just two identically situated identical agents. In the first case, the agents have no choices and are simply exposed to the

6. Dworkin, "What Is Equality? Part 2: Equality of Resources"; Eric Rakowski, *Equal Justice* (New York: Oxford University Press, 1991); Hillel Steiner, *An Essay on Rights* (Cambridge, Mass.: Blackwell, 1994); Philippe Van Parijs, *Real Freedom for All* (New York: Oxford University Press, 1995).

7. Arneson, "Equality and Equal Opportunity for Welfare" and "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare"; Cohen.

8. Amartya Sen, "Equality of What?" in *The Tanner Lectures on Human Values*, ed. Sterling McMurrin (Salt Lake City: University of Utah Press, 1980), vol. 1, pp. 353-69, and *Inequality Reexamined* (Cambridge, Mass.: Harvard University Press, 1992).

same natural lottery—exposure to lightning strikes, say. The outcomes of the lottery are clearly due to brute luck, since no choice is involved at all. In the second situation, each agent has two choices—to stop or to continue traveling during a storm—but their choices have no effect on the probabilities of being struck by lightning. Whatever choice they make, they are exposed to the same natural lottery as in the first situation. Each agent is fully aware of the outcomes of each choice (and in particular that the probability of being struck by lightning is not affected by what choice he or she makes). Each agent chooses (arbitrarily) to stop, and then one is struck by lightning, and the other is not. Given that the possibility and probability of being struck by lightning, given their choices, were reasonably foreseeable by the agents, the above view of brute luck would classify this as nonbrute luck (option luck). There seems, however, to be no reason to treat this case differently from the initial case. Both involve unavoidable inequalities in outcome luck. There was nothing the agents could have done to alter their exposure to the lightning strikes. Hence, the presence of fully informed choice—even along with background equality—is not sufficient to make an outcome a matter of option luck.⁹

A better (normatively more relevant) understanding of brute luck is in terms of avoidability:

Brute Luck as (Reasonable) Unavoidability: The occurrence of an event is due to brute luck for an agent if and only if the agent could not have (reasonably) avoided the possibility of its occurrence.

The notion of avoidability (like that of foreseeability) is to be understood as lifetime avoidability. That is, the issue is whether the agent could have—at some point in his or her life—avoided the possibility of the result. The mere fact that a result is unavoidable for the agent shortly before its occurrence does not make the result a matter of brute luck. For the situation in which the result is unavoidable (e.g., that the loan shark will break my legs) may itself have been avoidable (e.g., by my not borrowing money from her at an earlier time). For simplicity, my

9. An important issue that I shall not address is the following: with respect to avoidability (and, below, influenceability), is it the specific events that matter or the impact on advantage? For example, suppose that no matter what I do I will be struck by lightning and suffer a disadvantage of 10. If I make one choice I will be struck and lose the use of my left arm (disadvantage of 10, say). If I make a different choice, I will be struck by a different lightning bolt and lose some of the function of my right arm (disadvantage 10 as well). My disadvantage of 10 is not avoidable, but some of the particular events and outcomes are. I'm inclined to think that, for the characterization of brute luck, it is the impact on advantage that matters, but I shall not attempt to resolve this issue. The examples and discussion below can be recast if necessary.

examples and related discussion will assume that what is unavoidable in a situation was not earlier avoidable.

This conception of brute luck rightly sees no difference between the lightning strikes of the first situation and those of the second situation above (since in both cases there is nothing that the agent could do to eliminate the possibility of the strikes). Most authors (e.g., Arneson, Cohen, Rakowski, and Roemer) employ at least something like this conception.¹⁰ Indeed, it is present in Dworkin's original definition:

Option luck is a matter of how deliberate and calculated gambles turn out—whether someone gains or loses through accepting an isolated risk he or she should have anticipated and *might have declined*. Brute luck is a matter of how risks fall out that are not in that sense deliberate gambles.¹¹

There is, however, a question concerning how avoidability is to be understood. On a strict construal, if an agent can avoid exposure to the risk of being struck by lightning by always lying on the ground, then lightning strike outcomes are not due to brute luck. A problem with this view is that it does not take into account the reasonability of avoiding the result in question. Lying on the ground may sometimes be reasonable, but it is certainly unreasonable in many contexts. Where lying on the ground is unreasonable, the fact that by doing so one can avoid lightning strikes seems irrelevant.

On a looser construal, brute luck is taken to include all and only outcomes that are not reasonably avoidable by the agent. If, for example, lying on the ground was reasonable (e.g., because of warnings about stormy conditions), then failure to do so will make the results a matter of option luck. If it was not (e.g., because weather conditions were expected to be nonstormy), then failure to do so will not automatically make the results a matter of option luck.

Of course, this raises the question of what counts as reasonable avoidability. An outcome is reasonably avoidable just in case there is, for the agent, some reasonable choice that avoids the outcome. But what makes a choice reasonable? One view is that it is in the agent's best interest. Another view is that it is adequately (either in absolute terms or relative to the best choice) in the agent's interests. Of course,

10. These authors each emphasize something like the ability to control the occurrence of the event. See Arneson, "Equality and Equal Opportunity for Welfare" and "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare"; Cohen; Rakowski, pp. 73–74; and John Roemer, "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs* 22 (1993): 146–66, and *Equality of Opportunity* (Cambridge, Mass.: Harvard University Press, 1998), chap. 3.

11. Dworkin, "What Is Equality? Part 2: Equality of Resources," p. 293; emphasis added.

there are other possibilities as well. I shall not attempt to resolve the issue of how reasonability should be understood in this context. Instead, I shall note a general problem with appealing to reasonability.

No matter how exactly reasonableness is understood, appealing to reasonable avoidability faces, it seems, a structural problem. Suppose that an agent has two choices and that the outcomes are completely determined by the choices (no risk is involved). (The same point can be made when outcomes are only probabilistically determined by choices.) The agent may choose to pursue an education and live a wonderful life, or she may choose to become a drug addict and live a terrible life. The first option, let us stipulate, is a reasonable choice, whereas the second is not. Hence, the wonderful life is not reasonably avoidable (because it's the only reasonable option), whereas the terrible life is. Thus, if brute luck is understood as not reasonably avoidable, then the wonderful life outcome is brute luck, and the terrible life outcome is viewed as option luck. It seems intuitively strange, however, to think that the wonderful life outcome is a matter of brute luck. The agent had a fully free choice (we can stipulate), and if she chose the wonderful life, it would seem that the outcome is a matter of option luck.¹² More generally, where two identically situated agents make different choices, the differential impact of their choices is, it seems, option luck. Furthermore, it seems strange that the classification of wonderful life outcome would change (from brute luck to option luck) if an option were added that produced an equally wonderful, although different, life. In this case, the original wonderful life outcome would be reasonably avoidable (and hence deemed option luck), since the added option is also a reasonable choice.

Perhaps these implications are not as troubling as they first seem. Alternatively, perhaps there is some other way of capturing the notion of reasonable avoidability.¹³ I shall not pursue the matter further here, since ultimately I will argue against the normative significance of the distinction between brute and option luck.

12. Note that here, as is standard in the literature, option luck is understood as the complement of brute luck. Hence, even though no luck is involved in the examples (since choices determine the outcomes), certain outcomes (e.g., the terrible life) are classified as option luck. The point of the example, however, could also be made if choice did not determine outcomes and luck was involved.

13. Martin E. Sandbu has insightfully argued that brute luck is luck that is present in the least risky reasonable prospect. See his "On Dworkin's Brute Luck–Option Luck Distinction and the Consistency of Brute Luck Egalitarianism," working paper (Harvard University, 2001), available from the author upon request. For insightful discussions of reasonability conditions, see Kasper Lippert-Rasmussen, "Egalitarianism, Option Luck, and Responsibility," *Ethics* 111 (2001): 548–79; and Michael Otsuka, "Luck, Insurance, and Equality," *Ethics* (in press).

The characterization of brute luck as (reasonable) unavoidability has some further problems. One is that at best it is a gross simplification of a more adequate general idea of inability to influence. Inability to avoid is one way that an agent can be unable to influence outcomes, but it is not the only way. Suppose that an agent has a choice between a 99 percent chance of 100 and a 1 percent chance of zero, on the one hand, and a 1 percent chance of 100 and a 99 percent chance of zero, on the other. The agent is unable to avoid the possibility of each of the two outcomes, but she is able to influence the chance of their realization. The above characterization ignores how agents can influence outcomes by influencing their probabilities. A more general characterization is the following:

Brute Luck as (Reasonable) Inability to Influence: The occurrence of an event is due to brute luck for an agent if and only if the agent could not have (reasonably) influenced the possibility or probability of its occurrence.

This is a more adequate characterization of brute luck, since it recognizes that avoidability is but one way that outcomes may be influenceable by an agent. There is, however, a further problem. The "if and only if" clause in this characterization is problematic, for it requires that an outcome be classified as nonbrute luck as long as the agent could have even some very minimal influence on the outcome. For example, if an agent has a choice between a 50 percent chance of 100 and a 50 percent chance of zero, on the one hand, and a 50.00001 percent chance of 100 and a 49.99999 percent chance of zero, on the other, then the above condition holds that the result is entirely due to option luck (since she has some, extremely minimal, influence over probabilities). If, however, the two choices had identical payoffs and probabilities, then the result would be classified as brute luck (since she would have no influence over the outcome). It seems rather strange that a minute difference in probabilities should convert the entire outcome from brute luck to option luck.¹⁴ A more adequate characterization would replace "if and only if" with "to the extent that" and then have some method of apportioning outcomes between brute luck and option luck based on the degree of influence the agent had. It is not, however, clear that any

14. A similar problem arises where one can minimally influence the nature and advantage value of the outcomes (as opposed to probabilities). The "if and only if" clause requires that, where there is no influence, the outcomes be classified as due to brute luck, but, where there is even a small influence, the outcomes be classified as due to option luck.

plausible method of apportionment exists. Hence, this is an open question for brute luck egalitarians to address.¹⁵

There is yet one more complication related to the reasonableness of choice that we need to consider. So far, we have been discussing issues that arise even when agents are fully informed. The new issue concerns how incomplete or false beliefs should be reflected in the characterization of brute luck. Often agents are unable to influence events, the possibility of which they cannot foresee, but this is not always so. Some events may be unforeseeable but nonetheless influenceable. For example, an agent in the late 1970s may not have been able to foresee the possibility of catching AIDS from unprotected sex, but he could nonetheless have influenced the probability thereof by abstaining from unprotected sex. Influenceability (as well as avoidability) is, at least as I have been understanding it, objectively determined, whereas reasonable foreseeability is partly subjectively determined. Hence, they are distinct issues.

For simplicity, we shall focus on the initial beliefs of agents so as to be sure that the false or incomplete beliefs are clearly a matter of brute luck. Agents are owed compensation for such doxastic defects if others do not suffer from them. That, however, is not the issue here. The issue concerns whether the outcomes of choices of two identical agents with the same imperfect beliefs in identical situations can be due to differential brute luck even if the agents can influence the outcomes.

Suppose two identical agents face the same choice situation. The first choice yields 100, and the second choice yields zero, but, because of their unavoidably imperfect (and identical) beliefs, they each reasonably view their choices as between 60 and 60. Suppose that Smith (arbitrarily) makes the first choice and ends up with 100, and Jones (arbitrarily) makes the second choice and ends up with zero. Because Jones's result was avoidable (and, indeed, reasonably avoidable based on the objective payoffs), this result is classified as option luck. It seems strange, however, for brute luck egalitarianism to hold that the resulting inequality is just. Although the zero result was as matter of fact avoidable, Jones could not have reasonably foreseen the results of her choice. It seems strange to hold Jones accountable for results that she could not have reasonably foreseen. It also seems strange to hold that Smith is somehow entitled (because of it being option luck) to the 100-unit advantage, when she only foresaw the possibility of a 60-unit advantage.

Although I argued above that the mere fact that the possibility of an event was a foreseeable result of the agent's choices is not a sufficient

15. Skepticism about a categorical distinction between brute luck and option luck has been independently raised by Marc Fleurbaey in "Egalitarian Opportunities," *Law and Philosophy* 20 (2001): 499–530.

condition for the outcome to be option luck, the above example suggests that it may nonetheless be a necessary condition. This suggests the following characterization of brute luck, where an agent deliberately influences an outcome just in case he or she influences its occurrence and is aware that she has this influence:

Brute Luck as Not (Reasonably) Deliberately Influenceable: The occurrence (or nonoccurrence) of an event is due to brute luck for an agent to the extent that the agent could not have (reasonably) deliberately influenced the possibility or probability of its occurrence (nonoccurrence).¹⁶

In the above example—in which the first choice yields 100 and the second choice yields zero, but each of the identical and identically situated agents reasonably believes that each choice yields 60—this characterization has the following implications. The agent that chose the first option reaped 40 units in excess of his reasonable expectations (100 vs. 60). The agent that chose the second option reaped a 60-unit shortfall from her reasonable expectations (zero vs. 60). The excess and the shortfall are deemed brute luck (since they were not deliberately influenceable) and equalized. This requires a 50-unit transfer from the first agent to the second (20 units to share the good brute luck of the first agent and another 30 units to share in the bad brute luck of the second agent). Each thus ends up with 50 units.

This completes my examination of the characterization of brute luck. Unavoidability is at the core of the characterization of brute luck, but it seems plausible also to include (1) events for which the agent has no ability to influence the probability and (2) events for which the agent is unaware of his or her ability to influence the probability (because of false or incomplete beliefs). It also seems plausible to take brute luck to include events that are not reasonably subject to deliberate influence (even if strictly speaking they are so influenceable), but it's not clear how this can be done. The main problem for brute luck egalitarianism, however, is

16. Dworkin's original definition of option luck includes a deliberateness requirement: "an isolated risk he or she should have anticipated and might have declined" ("What Is Equality? Part 2: Equality of Resources," p. 293). Rakowski is one of the few commentators who, when reformulating Dworkin's distinction, explicitly recognizes both uninfluenceability (or at least avoidability) and unforeseeability as individually sufficient conditions for brute luck: "The distinction is therefore between, on the one hand, risks that people must ineluctably bear or that, though they could in principle have avoided running, they had no reason beforehand to associate with an activity in which they engaged, and, on the other hand, all other risks that people knowingly run or of which they should be aware" (pp. 73–74). The issue of foreseeability is also discussed insightfully by Lippert-Rasmussen, "Egalitarianism, Option Luck, and Responsibility"; by Otsuka; and by Thomas Christiano, "Comments on Elizabeth Anderson's 'What Is the Point of Equality?'" available on-line at <http://www.brown.edu/Departments/Philosophy/bears/9904chri.html>, 1999.

that any relevant distinction between brute luck and option luck will be a matter of degree. Small differences in deliberate influenceability should not convert the entire outcome from brute luck to option luck, or vice versa. Brute luck egalitarianism thus needs some measure of the degree of brute luck. It's not at all clear, however, how this might be done. Furthermore, it's not clear how brute luck egalitarianism should be understood if brute luck is a matter of degree. This problem is thus a fundamental problem confronting brute luck egalitarianism.

Below I shall develop a second, and more fundamental, attack on brute luck egalitarianism. I shall argue that, even if there is a plausible measure of the degree of brute luck, it is not relevant for justice. Instead, justice requires that initial opportunities for advantage be equalized. Before arguing for this claim, I shall first clarify the notion of initial opportunities and contrast it with brute luck equality.

INITIAL OPPORTUNITIES

However exactly brute luck is understood, it will include luck in initial opportunities for advantage (e.g., one's initial genetic endowment and one's initial social position) and some kinds of outcome luck (i.e., some kinds of luck in how things turn out; e.g., whether an uninfluenceable contingent possibility of a lighting strike is realized). Not included in brute luck are outcomes that are suitably related to choice—option luck. This includes both outcomes that are directly chosen in some appropriate sense (e.g., avoidable foreseen and certain outcomes of one's choices; e.g., being wet when one chooses not to open one's umbrella in the rain) and those that are risky outcomes that are suitably related to one's choices (e.g., winning or losing a lottery for which one purchased a ticket)—option outcome luck. Option luck, the complement of brute luck, thus (somewhat misleadingly) includes the direct results of choices when no luck is involved as well as option outcome luck.

Thus, we have four factors that jointly determine what outcomes are realized: brute luck in initial opportunities, brute outcome luck, choices, option outcome luck. To make the role of these four factors maximally clear, it may be helpful to have a diagrammatic display. The brute luck initial opportunities can be represented by a horizontal decision tree in which each branch represents an empirically possible way one's life can go. Along the decision tree, there are branches at various nodes. At some nodes—chance nodes—the path taken is determined by "acts of nature" or by the choices of others. At other nodes—choice nodes—the path taken is determined by the choice of the agent in question. Thus, initial opportunities are represented by a full decision tree, choices by choice nodes, outcome luck (which may be brute or option) by what happens at chance nodes, and outcomes by full branches (or paths). Writing the probabilities under the chance node

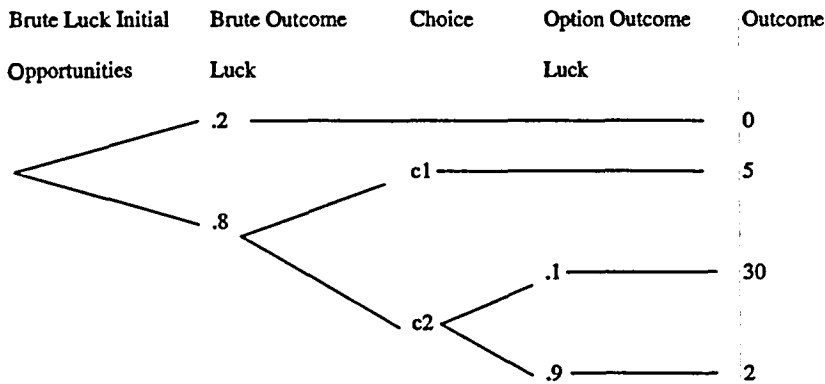


FIG. 1

columns, and the advantage value of each full branch under the outcome column, figure 1 below provides an example. This represents the initial opportunities of a perfectly informed agent, where there is a 20 percent chance of a brute luck event occurring that results in no further choices or chance events and has a value of zero. If this brute luck event does not occur, then the agent confronts the choice between *c1* and *c2*, both of which I'll assume are reasonable (their expected values are 5 and 4.8, respectively). If she chooses *c1*, then there are no further choices or chance events, and the outcome has a value of 5. If she chooses *c2*, then she is exposed to some option luck. She has a 10 percent chance of having a path worth 30 and a 90 percent chance of a path worth 2. (For simplicity, in this diagram brute outcome luck is shown as occurring prior to choice, but it may also occur after choices are made.)

Equality of initial opportunities for advantage calls for equality in the value of the initial decision trees that each agent confronts. Equality of brute luck calls for equality in the net result of the initial decision trees (initial brute luck) and (later) brute outcome luck events. These two views are each superior to two other egalitarian views. Equality of outcome advantage is problematic because it is not sensitive to the value people assign to having the freedom to engage in certain kinds of risky behavior (e.g., buying lottery tickets) and because it is highly insensitive to the administrative and incentive costs of outcome equalization (totally insensitive except where there is more than one way of equalizing outcome advantage). Some would also claim (although I shall argue against this view below) that it fails to hold people suitably accountable for their choices. Equality of luck holds that all factors except direct choice call for equalization. It calls for compensating for all kinds of luck (both brute and option luck) but allows uncompensated differences in out-

comes when they are due solely to differences in choices (e.g., when it is raining and no luck is involved, I knowingly choose to get wet by not using my umbrella and you knowingly choose not to be wet by using your umbrella). Although a slight improvement on equality of outcome, it still suffers from most of the main problems just mentioned. In particular, it implausibly always calls for the elimination of the effects of option outcome luck in chosen lotteries (e.g., regular lotteries).¹⁷

Equality of brute luck and equality of initial opportunities each hold that inequalities in initial opportunities are to be compensated. Brute luck equality also requires that inequalities in brute outcome luck be compensated. In the next section, I will argue against this second requirement.¹⁸ First, however, I'll briefly clarify two aspects of initial opportunity egalitarianism.

What point in a person's life counts as the "initial" point for the purposes of evaluating opportunities? There are two main candidates here—at the onset of (perhaps partial) moral standing and at the onset of some kind of full moral standing. If, for example, individuals acquire moral standing when they become sentient, but do not acquire a full set of moral rights until they are rational agents, then opportunities would be evaluated either at the onset of sentience or at the onset of agency. Arneson has tended to favor something like the full moral standing view (onset of adulthood), but this view is, I believe, implausible. For, if full moral standing is understood in absolute terms (e.g., as requiring rational agency), then sentient individuals not capable of rational agency (e.g., sufficiently mentally impaired individuals) will be left out of the equation. For the present purposes, however, we can leave this issue open.

How exactly are initial opportunities to be evaluated? Admittedly,

17. Equality of luck is at least roughly identical to what Arneson calls (misleadingly in my view, since it focuses on outcomes) "strict equality of opportunity." See his "Postscript (1995) to 'Equality and Equality of Opportunity for Welfare,'" in *Equality: Selected Readings*, ed. Louis Pojman and Robert Westmoreland (Oxford: Oxford University Press, 1997), pp. 238–42, and "Equality of Opportunity for Welfare Defended and Recanted." Equality of luck is not the view that Anderson calls "luck egalitarianism" in "What Is the Point of Equality?" This latter view is simply brute luck egalitarianism. It is misleading to call it "luck egalitarianism," since it does not call for equalizing option luck.

18. The fact that equality of initial opportunity does not guarantee equality of brute luck precisely because it does not guarantee equality of brute outcome luck was not initially adequately recognized. The divergence has also been noted—in the form of an objection to equality of initial opportunity—by Marc Fleurbaey, "Equal Opportunity for Equal Social Outcome," *Economics and Philosophy* 11 (1995): 25–55; and Kasper Lippert-Rasmussen, "Arneson on Equality of Opportunity for Welfare," *Journal of Political Philosophy* 7 (1999): 478–87. In his reply to the latter, Arneson acknowledges the divergence and (mistakenly, in my view) seeks to modify the equality of initial opportunity view so as to equalize for brute outcome luck. See his "Equality of Opportunity for Welfare Defended and Recanted."

this is a tricky issue, and there will probably be more than one plausible way of doing this, but there is no problem in principle. Here I shall identify some plausible approaches.

The first point to note is that we are concerned with effective opportunity for advantage. Hence, we must factor in not only the external opportunities but also the personal capacities and resources of the agent. As Arneson has emphasized,¹⁹ agents vary in their initial decision-making and decision-implementing abilities, and the evaluation of opportunity sets must adequately reflect this. A simpleminded idea would be to evaluate the opportunities on the assumption that an agent always chooses the best option of which he or she is capable. This will be an agent-relative matter, since the best feasible choice for one agent may not be feasible for another agent (e.g., due to a disability). This, however, is still not sufficiently sensitive to the agent's decision-making and implementing abilities. For some agents may be more prone to mistakes and weakness of the will, and this is not reflected if opportunity sets are evaluated on the basis of best choices. The best options available to each of two identical fallible agents may be equally good, but their second-best option might have radically different values. Given their fallibility, the difference in the second-best options matters.

Arneson proposes that two opportunity sets be judged equally valuable just in case (1) the best life path in the first opportunity set is equally valuable with the best life path in the second, (2) the second-best life path in the first opportunity set is equally valuable with the second-best life path in the second, and so on. This approach is plausible as a sufficient condition for opportunity sets being equally valuable. It fails, however, as a necessary condition. The best life path of one might be slightly worse than that of the second set, but, given that agents do not always choose their best option, this might be exactly offset by the second-best life path being better to an appropriate extent.²⁰

A more general and adequate approach will appeal to an agent's choice disposition, which, for each choice situation, specifies for each feasible option the probability of the agent choosing that option. For example, suppose that in a given choice situation there are just two options: D (having dessert), which will make the agent sick because she has already eaten too much, and \sim D (not having dessert), which is better for the agent. If the agent is perfectly rational, perfectly informed, not

19. Arneson, "Equality and Equal Opportunity for Welfare" and "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare."

20. It's possible, however, that Arneson is not giving a criterion of when opportunity sets are equally valuable but, rather, of when they are equal in the sense of having isomorphic payoff structures. This latter notion is stronger, and there is little reason, I believe, to require equality in this sense.

prone to mistakes, and so on, then the choice disposition will assign zero probability to the choice of D and 100 percent to the choice of $\sim D$. If the agent is somewhat disposed to suffer from weakness of the will, and so on, then her choice disposition might assign a 90 percent probability to the choice of $\sim D$ and a 10 percent probability to the choice D.²¹

Initial opportunity sets can thus be evaluated on the basis of their expected value. Probabilities are used for acts of nature, choices by others, as well as choices by the agent in question. Suppose, for example, that an agent faces an opportunity set with a 20 percent brute luck chance of ending with 100 (with no choices involved) and an 80 percent chance of having a choice. Suppose further that, if she has a choice, then she has a 90 percent disposition to make the first choice, which ends with 200, and a 10 percent disposition to make the second choice, which ends with zero. In this case, the expected value of the opportunity set is 164 ($= .2 \times 100 + .8 \times [.9 \times 200 + .1 \times 0]$). This effective opportunity set is more valuable than the opportunity set of a second agent who faces the same external opportunities but has a 20 percent (vs. 10 percent) choice disposition to make the second (less desirable) choice. The effective opportunity set of this second agent is, on the proposed method of evaluation, worth 148 ($= .2 \times 100 + .8 \times [.8 \times 200 + .2 \times 0]$). (In this artificial example, there is no issue of the earlier choices of agents later affecting their choice dispositions, but obviously the idea applies equally well when later choice dispositions can be influenced.)

One might question, of course, the theoretical legitimacy of ascribing probabilities to the choices of autonomous agents. This is indeed a deep and troubling issue, and all I can do is make a few hand-waving gestures. First, this approach need not invoke unique probability descriptions. It might instead invoke families of probabilities to reflect the choice-making dispositions of agents. At the limit, it might even reject the appeal to probabilities and simply appeal to the possible choices of an agent. Of course, the move to families of probability functions or to "possibility" functions will require some way of evaluating opportunity sets based on such functions, but well-developed approaches already

21. The idea of appealing to probabilistic choice dispositions is borrowed—with modification—from Roemer, "A Pragmatic Theory of Responsibility for the Egalitarian Planner" and *Equality of Opportunity*. He there appeals to the frequency distribution of how agents of the "same type" (having the same personal factors for which the agent is not responsible). I replace the frequency distribution with a choice disposition (which underlies the frequency distribution), since frequency distributions can fail to reflect to the underlying choice dispositions (especially when small numbers are involved). Roemer does not, however, evaluate opportunity sets on the basis of their expected value given the probability distribution. Instead, he invokes the frequency distributions of choices to impose a desert-based constraint on distribution of advantage.

exist for dealing with "vague probabilities," "probability interval ascriptions," and decision making under uncertainty (e.g., the maximin rule). Thus, such appeals need not create evaluation problems. More generally, however exactly it is done, it seems clear that there must be some way of ascribing dispositions to autonomous agents and having those dispositions affect the value of the effective opportunity sets.

There is, however, a problem with evaluating opportunity sets on the basis of expected value: it fails to allow for the possibility of rational risk aversion in this kind of life-encompassing situation. I used expected value merely as a way of illustrating how the idea could work. The more general idea leaves open (to be specified by the theory of personal advantage) how probabilities and possibilities are to be dealt with. The exact evaluation may build in some mild risk aversion, or even the radical risk aversion of maximin (maximize the value of the worst possible outcome). I leave all this open.

There is, however, a further generalization of the idea that is needed. For the above method of evaluation implicitly assumes that freedom (availability of choices) is only instrumentally valuable. This method holds that the values of outcomes determine—in conjunction with a risk and uncertainty rule—the value of opportunity sets. A more general approach, therefore, would not assume that the value of opportunity sets is reducible in this simple way to the value of outcomes. I take this to be a question that is to be answered by the theory of prudential advantage, which may allow that certain kinds of opportunities are intrinsically valuable.

In sum, although there are many different ways that the advantage value of opportunity sets can be assessed, and each way involves complex and difficult issues, there is no reason to suppose that a plausible theory of advantage cannot be developed so that opportunity sets can be evaluated. With the above understandings of brute luck and of initial opportunities, we are finally ready to ask which should be equalized.

JUSTICE DOES NOT REQUIRE THE EQUALIZATION OF BRUTE OUTCOME LUCK ADVANTAGE

Initial equality of opportunity for advantage is compatible with equality of brute luck advantage, but it does not require it. A scheme that provides equality of initial opportunities for advantage, but with no compensation for brute outcome luck (i.e., brute luck after the initial point), is one way of ensuring equality of initial opportunities, but it is not the only way. A second way is to provide (in addition) compensation for brute outcome luck. In both cases, initial opportunities for advantage are equal. I shall argue that justice requires compensation for brute outcome luck when and only when doing so is a way of increasing the value of people's initial opportunities.

Consider two schemes, each of which provides equality of initial opportunity for advantage, but only one of which provides compensation for brute outcome luck (e.g., acts of nature that are not deliberately influenceable). The scheme that provides compensation for brute outcome luck may have significantly higher administrative costs (since it needs to gather information about what happens, determine whether it is brute or option luck, and then collect and make payments). In addition, it may even have certain adverse incentive effects on people's behavior (e.g., it might for contingent reasons make people more prone to take various unreasonable risks). The net result of the administrative costs, incentives effects, and all other relevant factors may be that the initial opportunities for advantage faced by each person are less valuable under the brute outcome luck compensation scheme than without it. If that is so, then justice, I claim, forbids equalizing for brute outcome luck. For example, suppose (for simplicity) that no choices are involved, and that each person faces a 90 percent brute luck chance of 100 units and a 10 percent chance of zero units (expected value of 90). The only alternative is to provide compensation for the brute outcome luck, but, given very high administrative costs, and so on, this means that we provide only 10 units of advantage for each person (with brute outcome luck eliminated). It is implausible to hold that justice requires that brute outcome luck be equalized in this case. It would radically reduce the value (from 90 to 10) of everyone's initial opportunities.

The point is not that justice never permits compensation for brute outcome luck. It is rather that justice does not always require such compensation. Initial opportunity egalitarianism favors such equalization just to the extent that it efficiently promotes equality of initial opportunities (e.g., same degree of equality but with more valuable opportunity sets for all). To the extent that administrative costs are low, incentive effects are nonnegative, and the value of opportunity sets reflects some risk aversion, equality of initial opportunity for advantage will tend to favor compensating for bad brute outcome luck. The exact level of compensation provided for various kinds of brute luck, however, will vary depending on the costs and benefits of doing so. It deems it unjust, however, to provide compensation for brute luck where, for example, everyone would have equal and better life prospects without such compensation.

The superiority of initial opportunity egalitarianism to brute luck egalitarianism as a theory of justice follows from three claims—recalling that our topic is justice in the sense of giving people their due (as opposed to the merely comparative concern of fairness). The first claim is uncontroversial and holds that if we can give everyone more of what we owe them (noncomparatively) without upsetting the comparative balance (e.g., preserving equality among those who have an equal claim), then it

is unjust not do so. The second claim is that with respect to brute luck advantage of the relevant sort (to be specified below) individuals have a (noncomparative) claim to as much as possible compatible with whatever others get. This is a nonwaste condition. In conjunction with the first claim, it holds that it is unjust to give everyone one unit of the relevant brute luck advantage when everyone could be given two units. Any plausible egalitarian theory of justice will satisfy this condition. The third claim is that the relevant kind of brute luck advantage is initial opportunity for advantage (initial life prospects)—as opposed to all brute luck advantage (including brute outcome luck advantage). This, of course, is the controversial claim, and I will now defend it.

Both initial opportunity egalitarianism and brute luck egalitarianism are concerned with brute luck advantage. They take, however, two different perspectives on the matter. Initial opportunity egalitarianism takes an *ex ante* perspective (focusing on probabilities of brute luck advantage) and brute luck egalitarianism takes an *ex post* perspective (focusing on how brute luck advantage turns out). The latter perspective is highly insensitive to the costs (in terms of brute luck advantage) of achieving *ex post* equality, whereas the former perspective factors them in. If there are no net (e.g., administrative or incentive) costs in achieving *ex post* equality and if the value of opportunity sets is based on even slight risk aversion, then the two approaches agree that *ex post* equality of brute luck is required. If, however, there are costs to such equality, then the *ex ante* perspective factors them in and may not require *ex post* equality (e.g., when all can face a 90 percent chance of 100 and a 10 percent chance of zero or get 10 for sure). The *ex post* perspective, on the other hand, ignores these costs (and requires 10 for sure in the case just mentioned). The *ex ante* perspective is, I claim, more plausible. For it is the perspective that rational individuals would adopt if they were offered a choice (*ex ante*, of course) on how to deal with their own brute luck. Rational individuals would insure against some but not all possible brute outcome luck events. Given administrative costs, and so on, the insurance premiums for some of these events are just too expensive.

Of course, many (indeed probably most!) will be unconvinced by this willingness to leave brute outcome risks uncompensated. Let us therefore consider some specific objections.

Brian Barry asks us to consider a case in which there are systematically massive differences in life chances—a caste system say.²² This strikes us as

22. Brian Barry, "Equal Opportunity and Moral Arbitrariness," in *Equal Opportunity*, ed. Norman E. Bowie (Boulder, Colo.: Westview, 1988), pp. 23–44. The point also appears in Brian Barry, *Theories of Justice* (Berkeley: University of California Press, 1989), p. 224n. The idea of a lottery for unequal family positions is also discussed in James Fishkin, *Justice*,

quite unjust. Our view would, he suggests, be unaffected if we were to discover that at birth (conception, sentience, or adulthood) individuals are assigned to castes based on some equal opportunity randomization device. According to initial opportunity egalitarianism, however, this makes, it seems, all the difference in the world. Each person would have initial equality of opportunities for advantage in this case. Each person's caste is the result of brute outcome luck, and hence justice may not require compensation for this. Admittedly, Barry has commonsense intuition on his side.

To keep things simple, let us assume that the caste systems involved give all individuals the same rights of noninterference (e.g., of bodily integrity). For I fully agree that any caste system that does not respect certain rights of noninterference is unjust. Here, however, we are focusing on the demands of material equality, and so I want to set this issue aside. Let us suppose that the caste system gives everyone the same rights of noninterference but gives positive rights and opportunities that vary (and vary in value) by caste.

Should mere equality of opportunity be rejected as too weak given that it can judge caste systems as just? A first reply is that caste systems are typically inefficient as ways of promoting equality of initial opportunities for advantage (due to waste of human talent, resentment, etc.). Hence, caste systems will typically be judged unjust by the equality of opportunity view as well. Moreover, to the extent that the evaluation of opportunity sets reflects risk aversion, even if caste systems are efficient with respect to resources, they may not be efficient with respect to the value of opportunity sets (e.g., a 50 percent chance of 100 and a 50 percent chance of zero may be judged as less valuable than a 100 percent chance of 40). Indeed, given that I have remained neutral on the extent to which the evaluation of opportunity sets reflects risk aversion, nothing I have said rules out extreme risk aversion in the form of maximin. (In this case, initial opportunity egalitarianism and brute luck egalitarianism are effectively equivalent.) If opportunity sets are so evaluated, then caste systems will be ruled out except where they make the worst-off person as well-off as possible. Given, however, that extreme risk aversion is not a plausible way of evaluating opportunity sets, this reply has only limited force.

Of course, egalitarians concerned with equality of social status will be totally baffled by the above willingness to envisage the justness of caste systems.²³ For equality of social status is, indeed, totally lacking in the envisaged situation. Because I am here remaining neutral on the

Equal Opportunity, and the Family (New Haven, Conn.: Yale University Press, 1983), pp. 57–58, 61–64, and 110–13.

23. See Miller; Anderson; Wolff; and Mason.

relevant conception of outcome advantage, I will not rehearse the inadequacies of taking social status to the relevant conception (roughly: it fails to take account of all the other things that matter to individuals).²⁴ Instead, I will simply note that brute luck egalitarianism faces a comparable problem if it holds that option luck advantage need not be equalized. Unequal option luck can produce inequalities in social status just as much as unequal brute outcome luck can. Indeed, caste systems could arise merely as a matter of option luck. Of course, the inequalities of social status are more likely to arise on initial opportunity egalitarianism, but the point here is that equality of social status is a purely outcome-based concern and hence not one that will provide a wedge between brute luck and initial opportunity egalitarianism.

A second objection to initial opportunity egalitarianism (raised by one of the editors of this journal) is the following. Suppose that one group of people initially faces a 99 percent chance of a wonderful life and a 1 percent chance of a miserable life, whereas a second group of people faces the opposite risk (a 1 percent chance of a wonderful life and a 99 percent chance of a miserable life). According to initial opportunity egalitarianism, all the individuals of the first group—including those whose lives turn out miserably (!)—owe compensation to all the individuals of the second group—including those whose lives turn out wonderfully. This, however, seems crazy. Why would a person with a miserable life owe compensation to someone with a wonderful life?

Admittedly, there is something troubling about this, but it is not as significant as it might seem. First, even brute luck egalitarians agree that sometimes someone with a miserable life owes compensation to someone with a wonderful life. For if someone has wonderful brute luck and then loses everything because of bad option luck, then he owes—according to brute luck egalitarianism—compensation to a person with terrible brute luck but who has a wonderful life because of wonderful option luck. We must therefore abstract from claims of need here. Neither initial opportunity egalitarianism nor brute luck egalitarianism recognizes them. Because my goal here is only to defend the relative superiority of the former, I shall not here attempt to defend it against the criticism of insensitivity to needs.

There is, furthermore, a second reply to this objection. Initial opportunity egalitarianism will typically provide some compensation for brute outcome luck (e.g., where there is significant risk aversion and administrative costs and adverse incentive effects are not too great). Thus, typically, it may not require any compensation from those with better prospects who were unlucky and have miserable lives. It may

24. For elaboration on this point, see, e.g., Arneson, "Luck Egalitarianism and Prioritarianism."

instead require those with better prospects who were lucky and have wonderful lives to cover the entire cost of compensating those with worse prospects. Of course, there are possible situations in which initial opportunity egalitarianism will require those with good prospects but miserable lives to compensate those with bad prospects but adequate lives. Such situations, however, will be ones in which risk aversion and decreasing marginal advantage from resources are not significant enough to offset the administrative and (if any) incentive costs. In such situations, it is not so implausible to view this as just (assuming that issues of needs are set aside).

A related objection to the equality of initial opportunity for advantage is that it seems quite arbitrary to treat initial brute luck as calling for equalization but not to view brute luck a few seconds later as so calling. Why should disease that starts prior to the initial point (initial brute luck) call for equalization but not disease that strikes shortly after this point (brute outcome luck)?

A first reply is that initial opportunity egalitarianism does not draw the line as starkly as suggested. As already noted, it may well compensate for brute outcome luck when this is an efficient way of promoting equality of initial opportunities. Furthermore, it does not draw the line between what happens before the initial point and what happens after. Rather, it assesses the chances for advantage relative to that initial point. If at that point there is a 100 percent chance that a bad brute luck event will happen some twenty years later, the evaluation of the opportunity set will fully reflect the occurrence of that event. Thus, if one individual faces such an event and an otherwise identically situated agent does not, then initial opportunity egalitarianism requires that full compensation be given to the first individual. Furthermore, if the chance of the bad brute luck event is 99 percent, then it will require in principle almost full compensation.

Initial opportunity egalitarianism does not attach any principled significance to whether events occur before or after the initial evaluation point. It is, rather, that it evaluates the opportunity set relative to the chances involved at that point. The issue that separates brute luck and initial opportunity concerns compensation for how these chances turn out. Indeed, where the initial chance of an event is 100 percent (i.e., where it is fully deterministic), then both views agree that full compensation is owed. In this case, there is no relevant brute outcome luck. The event is simply a case of bad brute luck in initial opportunities.

A related objection to equality of initial opportunities is that it arbitrarily fails to provide for equalization for events that were completely unforeseeable at the beginning of the initial opportunity set. If, for example, equalization is provided between two children at the beginning of their lives, and it is later discovered that one of the children

had a previously undetected genetic predisposition to a certain disease, it seems quite unjust not to provide compensation simply because we were initially unable to detect this disposition.

The answer, of course, is that it would indeed be completely unjust to provide no later compensation in such a case. For as a matter of fact the initial opportunities were not equal. We believed that they were equal, but they were not. Hence, the discovery of the genetic disposition requires some additional compensation for the person in question.

This ends my defense of equality of initial opportunity for advantage and my criticism of equality of brute luck advantage. I now turn to a second issue that may seem to separate the two views: the treatment of option luck. Brute luck egalitarians have tended to condemn compensation for bad option luck. I shall argue, however, that egalitarians—including brute luck egalitarians—need not, and should not, always condemn such compensation.

JUSTICE PERMITS COMPENSATION FOR OPTION LUCK DISADVANTAGE

So far, we have focused on principles of equality (what should be equalized). A theory of justice may also posit principles of accountability (what should not be equalized). One way of equalizing brute luck advantage, for example, is by implementing a publicly announced and suitably proactive policy of equalizing outcome advantage. Most brute luck egalitarians, however, would reject this approach on the grounds that it equalizes the results of option luck, and these should not be equalized. They invoke a principle of accountability to rule out equality of outcome advantage. I shall argue against substantive principles of accountability. Neither brute luck egalitarianism nor equal initial opportunity egalitarianism need, nor should, endorse such principles. For simplicity, I will focus on brute luck egalitarianism, but the issue applies to all forms of egalitarianism.

Suppose that each agent starts with equally valuable opportunities for advantage, there is no brute outcome luck, and some individuals run cash lotteries (selling tickets and giving out cash prizes). These lotteries are reasonably avoidable and their possible results are fully foreseeable. Hence, any resulting inequalities are the results of option luck. Does justice permit the coercive imposition of a tax on lottery winnings to help those who are destitute because of losing all their money on lotteries? Below I shall address the question of whether justice would allow such taxation when the taxation scheme is introduced retroactively. First, however, I want to consider the public and proactive introduction of the scheme. That is, we shall consider the justice of such a taxation scheme where this scheme is publicly announced and

only applies to lottery winnings that take place after the public announcement.

Brute luck egalitarians have some tendency to hold that such a scheme is unjust. After all, without the scheme, the resulting inequalities would be the result of option luck. The idea is that, although coercive redistribution is called for to equalize brute luck advantage, it should not modify the "natural" option luck differential payoff structure. Instead, if there is full information, the redistribution scheme should simply tax each person an amount equal to his or her excess brute luck advantage and transfer to people an amount equal to their brute luck advantage shortfall. The basic option luck payoff structure should not be modified. Taxes should not be choice sensitive (e.g., you should not have to pay higher taxes if you chose to play the lottery and win).

Brute luck egalitarians, that is, have tended to endorse something like the following principle of accountability:

Natural Rewards: If, prior to any coercive transfers, two agents are identical with respect to the factors for which justice requires equalization, then justice requires that the two agents have the same coercive transfers (e.g., taxes or subsidies).²⁵

This is a generic principle; it leaves open what factors must be equalized. It says that whatever they are (e.g., brute luck advantage or initial opportunities for advantage), there should be no transfers beyond those required for the relevant equality. If, prior to the transfers, two agents are already equal in the relevant respects, then their transfers should be the same. Their net transfers (taxes or subsidies) should not depend on what choices they make. For brute luck egalitarians, this means that there should be no equalization that is not required for brute luck equalization.

This principle, however, is implausible. There is no reason to treat the pre-coercive-redistribution—or "natural"—payoff structure as privileged. Typically, there will be many redistributive schemes that will

25. Dworkin's condition of ambition-sensitivity for the justness of resource allocations—which requires that gains and losses from choices not be redistributed when made from an appropriate position of equality—is a version of Natural Rewards. See "What Is Equality? Part 2: Equality of Resources," p. 311. In the economics literature, something like Natural Rewards is known as Equal Resources for Equal Nonresponsible Characteristics (or some variant thereof), and talents and handicaps are typically assumed to be the nonresponsible characteristics. A weaker natural rewards principle requires only that there be no transfers when everyone has the same (or equally valuable) nonresponsible characteristics. My argument in the text applies against this weaker principle as well. For formulation and discussion of natural reward conditions, see, e.g., Fleurbaey, "Equal Opportunity for Equal Social Outcome," and "Equality among Responsible Individuals," in *Freedom in Economics: New Perspectives in Normative Analysis*, ed. Jean-François Laslier, Marc Fleurbaey, Nicolas Gravel, and Alain Trannoy (London: Routledge, 1998), pp. 206–34.

maximally equalize brute luck advantage. Leaving the natural (pretransfer) option luck differential payoff structure in place (and taxing away excess brute luck advantage) will be one of them, but it may not be the most efficient way of maximally promoting brute luck equality. A modification of the natural option luck payoff structure may, in conjunction with taxes on brute luck advantage, also maximally equalize brute advantage, and it may produce higher levels of brute luck advantage for everyone. This will be the case, for example, where risk aversion leads to an underproduction of goods. Let me explain.

Suppose that all agents start with equally valuable opportunities (initial brute luck) and that (for simplicity) there is no brute outcome luck. There is no need for coercive transfers for the purposes of achieving brute luck equality. Suppose that there are just two ways of providing health care services. One way is to have no coercive transfers for that purpose and to let individuals purchase individual insurance on the free market. This is the scheme favored by Natural Rewards (relative to brute luck egalitarianism as well as relative to initial opportunity egalitarianism), since it leaves "natural" option luck in place. The only alternative, let us suppose, is to coercively tax everyone an equal amount to fund the health services and then to provide all basic health services independently of whether the health problems are the result of brute luck (e.g., lung problems even if brought on solely because of smoking or liver problems induced solely by excessive drinking). This violates Natural Rewards, since the net transfers (taxes plus benefits) are not needed to achieve the requisite equality (which by stipulation was already present). Although the (equal per capita) taxes are choice insensitive, the funded benefits are not. Those who develop lung problems because they choose to smoke will, all else being equal, gain greater net funded benefits than those with the same brute luck (e.g., genetic disposition for lung cancer) who choose not to smoke.

There is, however, no principled reason to hold that coercive funding of universal health services is unjust. Both it and the market approach will equalize brute luck and be suitably proactive and publicly announced. If the administrative costs and adverse incentives of the coercively funded scheme are sufficiently high, then the market approach will be a more efficient way of equalizing brute luck advantage. If, however, the administrative costs of the coercively funded scheme are sufficiently low (or even just sufficiently lower than the market approach!), the incentive effects are not significantly adverse (or perhaps even positive), and people are sufficiently risk averse, then the coercively funded scheme may be more efficient. It may be, that is, that the coercively funded scheme will provide just as much equality, but with individuals at a higher level of brute luck advantage. If it is, then justice, I claim, requires the coercively funded scheme—even though some peo-

ple will not be fully bearing the costs of their "natural" option luck health decisions. It would be unjust, because inefficient, to insist—as Natural Rewards does—that there be no choice-sensitive transfers. Retroactive choice-sensitive transfers are indeed problematic, but public and proactive ones are not (since agents will take these transfers into account in their choices).

A more abstract example may help clarify the point. Here I will focus just on risk aversion and incentives and indicate how schemes that reduce the opportunity for option luck may be more efficient ways of equalizing brute luck advantage (or alternatively: the value of initial opportunities). I'll assume for simplicity that administrative costs are zero, and that the advantage equalisandum (e.g., well-being) is also what individuals rationally seek to promote. Suppose that water is in scarce supply and that each individual has a choice between exploring for water and not exploring. Exploration provides a 20 percent chance of failure and hence of receiving nothing and an 80 percent chance of success (finding water) and hence of receiving 10 units. Exploration thus has an expected value, for each individual, of 8 units. Nonexploration ensures, say, that one will receive 7 units of advantage for sure. Assume, as is reasonable, that the difference in outcomes is a matter of option luck. Because people are significantly risk averse, if this natural option luck is left in place, no one, let us suppose, would choose the risky option and everyone would end up with 7 units. Under a second scheme, however, full compensation is provided for those who choose the risky option and lose. Instead of getting nothing, they receive 8 units from a tax transfer scheme that is funded from the profits of those who choose the risky option and win. Under this scheme, the "risky" option ceases to be risky, since one will get 8 units for sure after transfers. Furthermore, this option is clearly preferable to the first option of 7 units for sure. Hence, everyone would choose the "risky" option. The result will thus be that 20 percent of the agents will lose and get nothing initially, and 80 percent of them will win and get 10 units initially. Then, as announced in advance, each of the winners is subject to a tax to provide 8 units to each of the losers. Each winner pays a tax of 2 ($= [8 \times .2]/.8$) and thus ends up with 8 units. Both schemes equalize brute luck advantage (since each ensures that the initial opportunities are the same for all and that there is no brute outcome luck) and are proactively public. Under the second scheme, however, each person's brute luck advantage is higher than it is under the first.

Natural Rewards holds that agents have a preinstitutional entitlement to reap their "natural rewards," where these are understood as the rewards that would follow if no coercive redistribution were implemented between agents who are identical with respect to the factors that must be equalized (e.g., brute luck or initial opportunities). Brute

luck egalitarians have been attracted to Natural Rewards, I think, because they view it as unjust for someone to be coercively required to subsidize someone else's option luck (e.g., the health cost of smoking, when this is genuine option luck).²⁶ This point, however, does not support Natural Rewards. People should indeed reap the benefits and bear burdens in accordance with the payoffs as defined by public and proactive just institutions. There is, however, no reason to suppose that just institutions must leave "natural" rewards in place. The natural reward structure is just one among many possibilities.

One might think that respecting the natural reward structure is required if the reward scheme is to be neutral among competing conceptions of the good life. Altering the "natural" reward structure, it may be thought, would favor some conceptions over others.²⁷ This, however, is not so. We are assuming throughout that the reward scheme must satisfy brute luck (or initial opportunity) egalitarianism. As long as the relevant conception of advantage is suitably neutral (e.g., some kind of preference satisfaction theory), there will be the relevant equality and neutrality. All will face equal brute luck advantage.

Natural Rewards is implausible because it views the "natural" option luck payoff structure as if it were normatively privileged, but it is not. Furthermore, by doing so, it is incompatible with our duty to others to provide them as much of the relevant equalisandum as possible (given what others are receiving). It rules out schemes that may be just as equalizing and better for everyone than schemes that it allows.²⁸ Thus,

26. For example, in *Contemporary Political Philosophy* (New York: Oxford, 1990), Will Kymlicka writes: "This points to an important component of our everyday sense of what it means to treat people as equals—namely, we should not expect others to subsidize our projects at the expense of their own" (p. 40). "Treating people with equal concern requires that people pay the cost of their own choices. . . . It is equally unjust for me to demand that someone else pay for the costs of my choices" (p. 75). For related discussions of this issue, see Terry Price, "Egalitarian Justice, Luck, and the Costs of Chosen Ends," *American Philosophical Quarterly* 36 (1999): 267–78; Mason; and Colin MacLeod, *Liberalism, Justice, and Markets* (Oxford: Oxford University Press, 1998).

27. See, e.g., pp. 210–11 of Fleurbaey, "Equality among Responsible Individuals," and pp. 685–86 of "Equality and Responsibility," *European Economic Review* 39 (1995): 683–89.

28. Marc Fleurbaey has noted the incompatibility of Natural Rewards with Pareto efficiency in contexts in which individuals are accountable for their preferences. My criticism appeals to incentive effects and risk aversion and does not depend on holding people accountable for their preferences (which is implausible when this is a matter of brute luck). See p. 212 of Fleurbaey, "Equality among Responsible Individuals," and pp. 34–36 of "Equal Opportunity Egalitarianism or Equal Social Outcome," *Economics and Philosophy* 11 (1995): 25–55.

brute luck egalitarians and initial opportunity egalitarians can and should reject Natural Rewards.²⁹

There is, however, a related principle that is plausible. Natural Rewards places a general constraint on the structure of payoffs. It opposes "unnecessary" transfers (i.e., transfers that are not necessary for equalizing the relevant factors) even when this is public and proactive. The related and more plausible principle has no problem with suitably public and proactive modifications of "natural" payoffs. It rejects, however, coercive redistributions that are not suitably public and proactive. Rawls is appealing to something like this principle when he writes, "It is perfectly true that given a just system of cooperation as a scheme of public rules and the expectations set up by it, those who, with the prospect of improving their condition, have done what the system announces that it will reward are entitled to their advantages. . . . But this sense of desert presupposes the existence of the cooperative scheme; it is irrelevant to the question of whether in the first place the scheme is to be designed in accordance with the difference principle or some other criterion."³⁰

Consider, then:

Institutional Rewards: If, prior to any coercive transfers, two agents are identical with respect to the factors for which justice requires equalization, then justice requires that they have transfers that have the same excess/shortfall relative to the transfers dictated by the public institutional norms governing transfers that were in effect shortly before their choices.³¹

Very crudely, the basic idea of this principle is that transfers should be done in accordance with institutional norms that are public (e.g., of

29. Frank Vandembroucke has, I have discovered from Marc Fleurbaey, independently made a similar criticism of Natural Rewards in chap. 1 of his *Social Justice and Individual Ethics in an Open Society: Equality, Responsibility, and Incentives* (Berlin: Springer, 2001).

30. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), p. 103. For additional discussion of legitimate institutional expectations see pp. 310–15 and pp. 188–89 of T. M. Scanlon, "The Significance of Choice," in *The Tanner Lectures on Human Values*, ed. Sterling McMurrin (Cambridge: Cambridge University Press, 1988), vol. 8, pp. 151–216. For an argument that a noninstitutional conception of rewards/desert is needed, see Samuel Scheffler, "Responsibility, Reactive Attitudes, and Liberalism in Philosophy and Practice," *Philosophy & Public Affairs* 21 (1992): 299–323. For related discussion, see Owen McLeod, "Desert and Institutions," in *What Do We Deserve? A Reader on Justice and Desert*, ed. Louis P. Pojman and Owen McLeod (New York: Oxford University Press, 1999), pp. 186–95.

31. For simplicity, I leave out one important qualification. The principle is to be understood as allowing deviations from legitimate expectations where this is the morally appropriate response to a past wrong. Legitimate expectations need not be respected where an independent wrong is involved. A similar qualification is needed for Natural Rewards.

which most people are aware) and in effect at the time of choice. The principle does not, however, require strict adherence to the public norms in place at the time of choice. For to do so would in general rule out the possibility of equalizing the relevant factors (e.g., brute luck)—since typically the public institutional norms in place do not ensure the requisite equality. The principle therefore requires something weaker than strict adherence to the public norms in place. It allows transfers to diverge from the requirements of the public norms but only as long as the divergence is the same—no matter what choices they have made—for agents that are identical with respect to the factors that must be equalized. Thus, although agents will not in general get what the public norms in place at the time of their choices required (since some adjustment will be necessary so as to equalize appropriately), the discrepancy will not depend on what choices they made.

Obviously, there are several murky issues lurking here. What exactly is required for a norm to be suitably public? How much advance notice is needed for proactivity? Some of these issues, no doubt, could be cleaned up by a more thorough analysis. There will also surely be some inevitable vagueness in these notions. The basic idea, I hope, should be clear enough.

Something like Institutional Rewards is plausible for egalitarians. Of course, those who endorse independent principles of preinstitutional desert or entitlement will have reason to reject this principle and impose their preferred principle of accountability.³² My point here is that there is no need for egalitarians to endorse such preinstitutional principles of accountability. They can endorse a robust principle of accountability simply by endorsing Institutional Rewards.

Once it is recognized that Natural Rewards—in addition to privileging the pretransfer distribution—has absolutely no sensitivity to the prevailing institutional norms governing transfers (and the expectations that they generate), its implausibility is obvious. As a retroactive principle, Natural Rewards' insensitivity to legitimate institutional expectations is fatal. As a proactive principle, it has no problem in that regard, but its arbitrariness is now obvious. Equality should be promoted in ways that ensure individuals reap the differential benefits and bear the differential burdens of their choices in accordance with the institutional norms publicly in place at the time of their choices, but robust accountability does not require that individuals get their "natural" rewards.

32. A defender of strong preinstitutional desert is George Sher, *Desert* (Princeton, N.J.: Princeton University Press, 1987). John Roemer and Richard Arneson each advocate theories of justice in which preinstitutional desert plays a role. See Roemer, "A Pragmatic Theory of Responsibility for the Egalitarian Planner" and *Equality of Opportunity*; and Arneson, "Equality" and "Equality of Opportunity for Welfare Defended and Recanted."

Before closing, we should note that once Natural Rewards is rejected, brute luck (and initial opportunity) egalitarianism is less prone to two objections based on its alleged treatment of option luck. First, it has been argued that brute luck egalitarianism can be unduly harsh in its noncompensation of extremely bad option luck.³³ We can now see that this does not follow as immediately as claimed—at least where public norms for the future are at issue. It may take a harsh line and provide no compensation for bad “natural” option luck. It will do this, however, only where brute luck equality is more efficiently promoted (e.g., due to administrative costs and adverse incentive effects). In many cases (e.g., where there are favorable incentive effects and people are extremely risk averse), it will provide some partial (or perhaps full) compensation for bad “natural” option luck. Of course, this does not eliminate the objection, since efficiency may well dictate taking a harsh line. It does, however, soften the objection, since in such cases the harsh line will have a plausible rationale.

It has also been objected that invoking the distinction between brute and option luck can be highly intrusive and demeaning.³⁴ It requires highly specific information about individuals and classifies some as inferior in the sense of having a lower capacity for a good life. This may well be so, but brute luck egalitarianism can be sensitive to these costs. To the extent they are present, there is a reason not to base the institutional norms on the distinction between natural brute and option luck. If these costs (effects on people’s outcome advantage) are great, it may favor providing compensation based on need or simply universal provision. Once the commitment to not equalizing factors that need not be equalized is dropped, it’s an open question what kind of option luck compensation scheme brute luck egalitarianism will favor. Again, this does not eliminate the objection, but it does soften it.

In summary, egalitarians generally (including initial opportunity egalitarians), and brute luck egalitarians in particular, should reject Natural Rewards and instead base their principles of accountability on Institutional Rewards.

CONCLUSION

The distinction between brute and option luck is not, I have argued, as clear as is generally supposed. The core notion for brute luck is that of unavoidability, but this needs to be generalized to the absence of (strict or reasonable) deliberate influenceability. The crucial outstanding problems for brute luck egalitarianism are (1) identifying some

33. For example, Fleurbaey, “Equal Opportunity for Equal Social Outcome”; and Anderson.

34. Wolff; Anderson.

relevant manner of apportioning degrees of influenceability so that the presence of a trivial amount of deliberate influenceability does not convert everything to option luck and (2) identifying how brute luck egalitarianism is to be understood once the distinction between brute and option luck is understood as a matter of degree.

I have further argued that justice does not, in any case, require that brute luck advantage be equalized, nor that natural option luck advantage be left in place. Justice requires that initial opportunities for advantage be equalized, but this need not (although it may) compensate for inequalities in brute outcome luck, and it may favor compensating for inequalities in natural option luck. The extent to which such compensation should be provided is a contingent matter and is determined by how efficiently it promotes equality of initial opportunities for advantage. We owe this efficiency to others.